PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68

A2

(11) International Publication Number: WO 99/57318

(43) International Publication Date: 11 November 1999 (11.11.99)

(21) International Application Number: PCT/US99/10251

(22) International Filing Date: 7 May 1999 (07.05.99)

(30) Priority Data:
09/074,936 7 May 1998 (07.05.98) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 09/074,936 (CON)
Filed on 7 May 1998 (07.05.98)

(71) Applicant *(for all designated States except US)*: SEQUENOM, INC. [US/US]; 11555 Sorrento Valley Road, San Diego, CA 92121 (US).

(71)(72) Applicant and Inventor: HILLENKAMP, Franz [DE/DE]; Bahlmannstrasse 5, D–48147 Münster (DE).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: KÖSTER, Hubert [DE/US]; 8636–C Via Mallorca Drive, La Jolla, CA 92037 (US).

(74) Agent: SEIDMAN, Stephanie, L.; Heller Ehrman White & McAuliffe, Suite 700, 4250 Executive Square, La Jolla, CA 92037 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: INFRARED MATRIX–ASSISTED LASER DESORPTION/IONIZATION MASS SPECTROMETRIC ANALYSIS OF MACROMOLECULES

(57) Abstract

Mixtures containing a biological macromolecule, such as a nucleic acid molecule or a polypeptide, and a liquid matrix, which absorbs infrared (IR) radiation, are provided. These mixtures are useful for analysis of the biological macromolecule by IR matrix assisted laser desorption/ionization (IR–MALDI) mass spectrometry. Also provided are processes for analyzing a biological macromolecule using IR–MALDI mass spectrometry. For example, processes for detecting the presence or identity of a biological macromolecule in a sample, or for sequencing a biological macromolecule are provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INFRARED MATRIX-ASSISTED LASER DESORPTION/IONIZATION MASS SPECTROMETRIC ANALYSIS OF MACROMOLECULES

RELATED APPLICATIONS

For U.S. purposes, this application is a continuation-in-part of U.S. application Serial No. 09/074,936, filed May 7, 1998, to Franz Hillenkamp, entitled "IR-MALDI Mass Spectrometry of Nucleic Acids Using Liquid Matrices." Where permitted the subject matter this application is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

The disclosed processes relate generally to the field of genomics, proteomics and molecular medicine, and more specifically to processes of using infrared matrix assisted laser desorption-ionization mass spectrometry to analyze, or otherwise detect the presence of or determine the identity of a biological macromolecule.

BACKGROUND OF THE INVENTION

In recent years, the molecular biology of a number of human genetic diseases has been elucidated by the application of recombinant DNA technology. More than 3000 diseases are known to be of genetic origin (Cooper and Krawczak, "Human Genome Mutations" (BIOS Publ. 1993)), including, for example, hemophilias, thalassemias, Duchenne muscular dystrophy, Huntington's disease, Alzheimer's disease and cystic fibrosis, as well as various cancers such as breast cancer. In addition to mutated genes that result in genetic disease, certain birth defects are the result of chromosomal abnormalities, including, for example, trisomy 21 (Down's syndrome), trisomy 13 (Patau syndrome), trisomy 18 (Edward's syndrome), monosomy X (Turner's syndrome) and other sex chromosome aneuploidies such as Klinefelter's syndrome (XXY).

Other genetic diseases are caused by an abnormal number of trinucleotide repeats in a gene. These diseases include Huntington's disease, prostate cancer, spinal cerebellar ataxia 1 (SCA-1), Fragile X syndrome (Kremer et al., Science 252:1711-14 (1991); Fu et al., Cell 67:1047-58 (1991); Hirst et al., J. Med. Genet. 28:824-29 (1991)); myotonic dystrophy type I

(Mahadevan et al., Science 255:1253-55 (1992); Brook et al., Cell 68:799-808 (1992)), Kennedy's disease (also termed spinal and bulbar muscular atrophy (La Spada et al., Nature 352:77-79 (1991)), Machado-Joseph disease, and dentatorubral and pallidolyusian atrophy. The aberrant number of triplet repeats can be located in any region of a gene, including a coding region, a non-coding region of an exon, an intron, or a regulatory element such as a promoter. In certain of these diseases, for example, prostate cancer, the number of triplet repeats is positively correlated with prognosis of the disease.

Evidence indicates that amplification of a trinucleotide repeat is involved in the molecular pathology in each of the disorders listed above. Although some of these trinucleotide repeats appear to be in non-coding DNA, they clearly are involved with perturbations of genomic regions that ultimately affect gene expression. Perturbations of various dinucleotide and trinucleotide repeats resulting from somatic mutation in tumor cells also can affect gene expression or gene regulation.

Additional evidence indicates that certain DNA sequences predispose an individual to a number of other diseases, including diabetes, arteriosclerosis, obesity, various autoimmune diseases and cancers such as colorectal, breast, ovarian and lung cancer. Knowledge of the genetic lesion causing or contributing to a genetic disease allows one to predict whether a person has or is at risk of developing the disease or condition and also, at least in some cases, to determine the prognosis of the disease.

Numerous genes have polymorphic regions. Since individuals have any one of several allelic variants of a polymorphic region, each can be identified based on the type of allelic variants of polymorphic regions of genes. Such identification can be used, for example, for forensic purposes. In other situations, it is crucial to know the identity of allelic variants in an individual. For example, allelic differences in certain genes such as the major histocompatibility complex (MHC) genes are involved in graft rejection or graft versus host disease in bone marrow transplantation. Accordingly, it is highly desirable to develop rapid, sensitive, and accurate methods for determining the identity of allelic variants of polymorphic regions of genes or genetic lesions.

Several methods are used for identifying allelic variants or genetic lesions. For example, the identity of an allelic variant or the presence of a genetic lesion can be determined by comparing the mobility of an amplified nucleic acid fragment with a known standard by gel electrophoresis, or by hybridization with a probe that is complementary to the sequence to be identified. Identification only can be accomplished, however, if the nucleic acid fragment is labeled with a sensitive reporter function, for example, a radioactive (^{32}P, ^{35}S), fluorescent or chemiluminescent reporter. Radioactive labels can be hazardous and the signals they produce can decay substantially over time. Non-radioactive labels such as fluorescent labels can suffer from a lack of sensitivity and fading of the signal when high intensity lasers are used. Additionally, labeling, electrophoresis and subsequent detection are laborious, time-consuming and error-prone procedures. Electrophoresis is particularly error-prone, since the size or the molecular weight of the nucleic acid cannot be correlated directly to its mobility in the gel matrix because sequence specific effects, secondary structures and interactions with the gel matrix cause artifacts in its migration through the gel.

Applications of mass spectrometry in the biosciences have been reported (see Meth. Enzymol., Vol. 193, *Mass Spectrometry* (McCloskey, ed.; Academic Press, NY 1990); McLaffery et al., Acc. Chem. Res. 27:297-386 (1994); Chait and Kent, Science 257:1885-1894 (1992); Siuzdak, Proc. Natl. Acad. Sci., USA 91:11290-11297 (1994)), including methods for mass spectrometric analysis of biopolymers (see Hillenkamp et al. (1991) Anal. Chem. 63:1193A-1202A) and for producing and analyzing biopolymer ladders (see, International Publ. WO 96/36732; U.S. Patent No. 5,792,664).

Mass spectrometry has been used for the analysis of nucleic acids (see, for example, Schram, Mass Spectrometry of Nucleic Acid Components, Biomedical Applications of Mass Spectrometry 34:203-287 (1990); Crain, Mass Spectrom. Rev. 9:505-554 (1990); Murray, J. Mass Spectrom. Rev. 31:1203 (1996); Nordhoff et al., Mass Spectrom. Rev. 15:67-138 (1997); U.S. Patent No. 5,547,835; U.S. Patent No. 5,605,798; PCT Application Publication No. WO94/16101; PCT Application Publication No. WO 96/29431).

The so-called "soft ionization" mass spectrometric methods, including Matrix-Assisted Laser Desorption/Ionization (MALDI) and ElectroSpray Ionization (ESI), allow intact ionization, detection and mass determination of large molecules, i.e., well exceeding 300 kDa in mass (Fenn et al., Science 246:64-71 (1989); Karas and Hillenkamp, Anal. Chem. 60:2299-3001 (1988)). MALDI mass spectrometry (MALDI-MS; reviewed in Nordhoff et al., Mass Spectrom. Rev. 15:67-138 (1997)) and ESI-MS have been used to analyze nucleic acids. Nucleic acids are very polar biomolecules that are difficult to volatize and, therefore, there has been an upper mass limit for clear and accurate resolution.

ESI has been used for the intact desorption of large nucleic acids even in the megaDalton mass range (Ferstenau and Benner, Rapid Commun. Mass Spectrom. 9:1528-1538 (1995); Chen et al., Anal. Chem. 67:1159-1163 (1995)). Mass assignment using ESI is very poor and only possible with an uncertainty of about 10%. The largest nucleic acids that have been accurately mass determined by ESI-MS are a 114 base pair double stranded PCR product (Muddiman et al., Anal. Chem. 68:3705-3712 (1996)) of about 65 kDA in mass and a 120 nucleotide E. coli 5S rRNA of about 39 kDa in mass (Limbach et al., J. Am. Soc. Mass Spectrom. 6:27-39 (1995)). Furthermore, ESI requires extensive sample purification.

MALDI-MS requires incorporation of the macromolecule to be analyzed in a matrix, and has been performed on polypeptides and on nucleic acids mixed in a solid (i.e., crystalline) matrix. In these methods, a laser is used to strike the biopolymer/matrix mixture, which is crystallized on a probe tip, thereby effecting desorption and ionization of the biopolymer. In addition, MALDI-MS has been performed on polypeptides using the water of hydration (i.e., ice) or glycerol as a matrix. When the water of hydration was used as a matrix, it was necessary to first lyophilize or air dry the protein prior to performing MALDI-MS (Berkenkamp et al. (1996) Proc. Natl. Acad. Sci. USA 93:7003-7007). The upper mass limit for this method was reported to be 30 kDa with limited sensitivity (i.e., at least 10 pmol of protein was required). Infrared MALDI-MS of proteins reportedly consumes 100-1000 times more material per spectrum as compared to UV MALDI-MS and, in combination with matrices such as glycerol,

can tend to form adducts which broaden the peaks on the high mass side (Hillenkamp et al. (1995) 43rd ASMS Conference on Mass Spectrometry and Allied Topics, p. 357). Furthermore, although IR-MALDI MS appeared to provide increased mass resolution due to less metastable fragmentation as compared to UV-MALDI MS, this decrease in metastable decay has been reported to be accompanied by an increase in fragmentation.

UV-MALDI-MS is limited in the size of biological macromolecules that can be analyzed. For example, it is difficult to analyze nucleic acid molecules much larger than about 100 nucleotides (100-mer) by UV-MALDI-MS.

Accordingly, despite the effort to apply mass spectrometry methods to the analysis of nucleic acid molecules, limitations remain due, in part, to physical and chemical properties of nucleic acids. For example, the polar nature of nucleic acid biopolymers makes them difficult to volatilize.

Analysis of large DNA molecules using UV-MALDI-MS has been reported (Ross and Belgrader, Anal. Chem. 69:3966-3972 (1997); Tang et al., Rapid Commun. Mass Spectrom. 8:727-730 (1994); Bai et al., Rapid Commun. Mass Spectrom. 9:1172-1176 (1995); Liu et al., Anal. Chem. 67:3482-3490 (1995); Siegert et al., Anal.Biochem. 243:55-65 (1997)). Based on these reports, it is clear that analysis of nucleic acids exceeding 30 kDa in mass (approximately a 100-mer) by UV-MALDI-MS becomes increasingly difficult with a current upper mass limit of about 90 kDa (Ross and Belgrader, Anal. Chem. 69:3966-3972 (1997)). The inferior quality of the DNA UV-MALDI spectra has been attributed to a combination of ion fragmentation and multiple salt formation of the phosphate backbone. Since RNA is considerably more stable than DNA under UV-MALDI conditions, the accessible mass range for RNA is up to about 150 kDa (Kirpekar et al., Nucl. Acids Res. 22:3866-3870 (1994)).

Nucleic acids in solid matrices (mostly succinic acid and, to a lesser extent, urea and nicotinic acid) have been analyzed by IR-MALDI (Nordhoff et al., Rapid Commun. Mass Spectrom. 6:771-776 (1992); Nordhoff et al., Nucl. Acids Res. 21: 3347-3357 (1993); Nordhoff et al., J. Mass Spec. 30:99-112 (1995)). Nordhoff et al. (1992) initially reported that a 20-mer of DNA and an 80-mer of RNA were about the uppermost limit for resolution. Nordhoff et al.

(1993) later provided distinct spectra for a 26-mer of DNA and a 104-mer of tRNA and reported that reproducible signals were obtained for RNA up to 142 nucleotides. Nordhoff et al. (1995) also reported a substantially better spectra for the analysis of a 40-mer by UV-MALDI with the solid matrix, 3-hydroxy picolinic acid, than by IR-MALDI with succinic acid, but that IR-MALDI resulted in a substantial degree of prompt fragmentation.

Analysis of macromolecules in a biological sample, for example, can provide information as to the condition of the individual from which the sample was obtained. For example, nucleic acid analysis of a biological sample obtained from an individual can be useful for diagnosing the existence of a genetic disease or chromosomal abnormality, a predisposition to a disease or condition, or an infection by a pathogenic organism, or can provide information relating to identity, heredity or compatibility. Since mass spectrometry can be performed relatively quickly and is amenable to automation, improved methods for obtaining accurate mass spectra for biological macromolecules, particularly for larger nucleic acid molecules larger than about 90 kDa for DNA and 150 kDA for RNA are needed.

Accordingly, a need exists for methods to detect and characterize biological macromolecules such as nucleic acid molecules, including methods to detect genetic lesions in a nucleic acid molecule. There is a need for accurate, sensitive, precise and reliable methods for detecting and characterizing biological macromolecules, particularly in connection with the diagnosis of conditions, diseases and disorders. Therefore it is an object herein to provide processes that satisfy these needs and provide additional advantages.

SUMMARY OF THE INVENTION

Processes for the determination of the mass or identity of biological macromolecules using infrared matrix assisted laser desorption/ionization (IR-MALDI) mass spectrometry and a liquid matrix are provided. In particular, infrared matrix assisted laser desorption/ionization (IR-MALDI) mass spectrometry of nucleic acids, including DNA and RNA, in a liquid matrix are provided. The liquid matrix (liquid at room temperature, one atmosphere pressure) is an IR-absorbing biocompatible material, such as a polyglycol,

particularly glycerol, that can form a glass or vitreous solid. The use of IR-MALDI and this liquid matrix can be employed in any method, particularly diagnostic methods and sequencing methods, heretofor performed with UV-MALDI. Such methods, particularly diagnostic methods for nucleic acids and proteins, include, but are not limited to, those described in U.S. Patent Nos. 5,547,835, 5,691,141, 5,605,798, 5,622,824, 5,777,324, 5,830,655, 5,700,642, allowed U.S. application Serial Nos. 08/617,256, 08/746,036, 08/744,481, 08/744,590, 08/647,368, published International PCT application Nos. WO 96/29431, WO 99/12040, WO 98/20019, WO 98/20166, WO 98/20020, WO 97/37041, WO 99/14375, WO 97/42348, WO 98/54751 and WO 98/26095.

In practicing an embodiment of the method for nucleic acid analyses, a composition for IR-MALDI containing the nucleic acid and a liquid matrix is deposited onto a substrate, which, generally, is a solid support, to form a homogeneous, transparent thin layer of nucleic acid mixture. This mixture is illuminated with infrared radiation so that the nucleic acid solution is desorbed and ionized, thereby emitting ion particles, which are analyzed using a mass analyzer to determine the mass of the nucleic acid. Preferably, sample preparation and deposition are performed using an automated device.

Methods for detecting the presence or absence of a biological macromoleculein a sample using IR-MALDI mass spectrometry are also provided herein. In a particular embodiment, a composition for IR-MALDI containing the biological macromolecule and a matrix is illuminated with infrared radiation, desorbed and ionized, thereby emitting ion particles, which are analyzed to determine whether the nucleic acid is present.

Methods for detecting the presence or absence of a nucleic acid in a sample using IR-MALDI mass spectrometry are also provided herein. In a particular embodiment, a composition for IR-MALDI containing the sample and a liquid matrix is illuminated with infrared radiation, desorbed and ionized, thereby emitting ion particles which are analyzed to determine whether the nucleic acid is present.

Liquid matrices for use in the processes disclosed herein have a sufficient absorption at the wavelength of the laser to be used in performing desorption and ionization and are a liquid at room temperature (20°C) and can form a vitreous or glass solid. The liquid is intended to be used in any IR MALDI format and at any temperature, typically about -200° C to 80° C, preferably -60° C to about 40° C, suitable for such formats.

For absorption purposes, the liquid matrix can contain at least one chromophore or functional group that strongly absorbs infrared radiation. Preferred functional groups include nitro, sulfonyl, sulfonic acid, sulfonamide, nitrile or cyanide, carbonyl, aldehyde, carboxylic acid, amide, ester, anhydride, ketone, amine, hydroxyl, aromatic rings, dienes and other conjugated systems.

Among the preferred liquid matrices are substituted or unsubstituted (1) alcohols, including glycerol, sugars, polysaccharides, 1,2-propanediol, 1,3-propanediol, 1,2-butanediol, 1,3-butanediol, 1,4-butanediol and triethanolamine; (2) carboxylic acids, including formic acid, lactic acid, acetic acid, propionic acid, butanoic acid, pentanoic acid and hexanoic acid, or esters thereof; (3) primary or secondary amides, including acetamide, propanamide, butanamide, pentanamide and hexanamide, whether branched or unbranched; (4) primary or secondary amines, including propylamine, butylamine, pentylamine, hexylamine, heptylamine, diethylamine and dipropylamine; and (5) nitriles, hydrazine and hydrazide. The liquids do not crystallize, but rather can form a glass or vitreous phase when subjected to drying, cooling or other conditions leading to a transition from the liquid phase. Materials of relatively low volatility are preferred to avoid rapid evaporation under conditions of vacuum during the IR-MALDI processes.

Preferably, a liquid matrix for use herein is miscible with a nucleic acid compatible solvent. As noted, it is also preferable that the liquid matrix is vacuum stable, i.e., has a low vapor pressure, so that the sample does not evaporate quickly in the mass analyzer. Preferably the liquid has an appropriate viscosity to facilitate dispensing of microliter to nanoliter volumes of matrix, either alone or mixed with a nucleic acid compatible solvent. Mixtures of different liquid matrices and additives to such matrices may be desirable to

confer one or more of the properties described above. Such mixtures can contain two liquid matrice materials (i.e.,. binary mixtures), three (tertiary mixtures) or more.

A nucleic acid/matrix composition for IR-MALDI is deposited as a thin layer on a substrate, which preferably is contained with a vacuum chamber. Preferred substrates for holding the nucleic acid/matrix solution can be solid supports, for example, beads, capillaries, flat supports, pins or wafers, with or without filter plates. Preferably the temperature of the substrate can be regulated to cool the nucleic acid/matrix composition to a temperature that is below room temperature.

Preferred infrared radiation is in the mid-IR wavelength region from about 2.5 μm to about 12 μm. Particularly preferred sources of radiation include CO, CO_2 and Er lasers. In certain embodiments, the laser can be an optic fiber laser, or the laser radiation can be coupled to the mass spectrometer by fiber optics.

In a further preferred embodiment, the ion particles generated by infrared irradiation of the analyte in the liquid matrix are extracted for analysis by the mass analyzer in a delayed fashion prior to separation and detection in a mass analyzer. Preferred separation formats include linear or reflector, with linear and nonlinear fields, for example, curved field reflectron; time-of-flight (TOF); single or multiple quadrupole; single or multiple magnetic sector; Fourier transform ion cyclotron resonance (FTICR); or ion trap mass spectrometers.

Processes of using IR-MALDI mass spectrometry to identify the presence of a target nucleic acid in a biological sample are provided. Such a process can be performed, for example, by amplifying nucleic acid molecules in the biological sample; contacting the amplified nucleic acid molecules with a detector oligonucleotide, which can hybridize to a target nucleic acid sequence present among the amplified nucleic acid molecules; preparing a composition for IR-MALDI, by mixing the product of the reaction with a liquid matrix, which absorbs infrared radiation; and identifying duplex nucleic acid molecules in the composition by IR-MALDI mass spectrometry, wherein the presence of duplex nucleic acid molecules identifies the presence of the target nucleic acid in the biological sample.

A process for identifying the presence of a target nucleic acid sequence in a biological sample also can be performed by amplifying nucleic acid molecules obtained from a biological sample; specifically digesting the amplified nucleic acid molecules using at least one appropriate nuclease, to produce digested fragments; hybridizing the digested fragments with complementary capture nucleic acid sequences, which are immobilized on a solid support and can hybridize to a digested fragment of a target nucleic acid to produce immobilized fragments; preparing a composition for IR-MALDI, containing the immobilized fragments and a liquid matrix, which absorbs infrared radiation; and identifying immobilized fragments by IR-MALDI mass spectrometry, thereby detecting the presence of the target nucleic acid sequence in the biological sample.

The presence of a target nucleic acid in a biological sample also can be identified by performing on nucleic acid molecules obtained from the biological sample, a first polymerase chain reaction using a first set of primers, which are capable of amplifying a portion of the nucleic acid containing the target nucleic acid; preparing a composition containing the first amplification product and a liquid matrix, which absorbs infrared radiation; and detecting the first amplification product in the composition by IR-MALDI mass spectrometry, thereby detecting the presence of the target nucleic acid in the biological sample. If desired, such a process can include, prior to preparing the composition for IR-MALDI, performing a second polymerase chain reaction on the first amplification product using a second set of primers that can amplify at least a portion of the first amplification product containing the target nucleic acid.

Also disclosed herein are compositions, particularly compositions for IR-MALDI, such compositions containing a biological macromolecule, which is suitable for analysis by IR-MALDI, and a liquid matrix, which absorbs infrared radiation. A biological macromolecule suitable for analysis by IR-MALDI can be, for example, a nucleic acid, a polypeptide or a carbohydrate, or can be a macromolecular complex such as a nucleoprotein complex, protein-protein complex, or the like. A composition for IR-MALDI as disclosed herein generally

contains the biological macromolecule, for example, a nucleic acid, and the liquid matrix in a ratio of about 10^{-4} to 10^{-9}, and can contain less than about 10 picomoles of biological macromolecule to be analyzed, for example, about 100 attomol to about 1 picomole (pmol) of the biological macromolecule. (For proteins, the analyte to matrix ratio is typicallyl narrower ranging fromabout 2 x 10^{-4} to 2 x 10^{-5}). A composition for IR-MALDI as disclosed herein also can contain an additive, which facilitates detection of the biological macromolecule by IR-MALDI, for example, an additive that improves the miscibility of the biological macromolecule in the liquid matrix. In one embodiment, a composition for IR-MALDI is deposited on a substrate, which can be a solid support such as a silicon wafer or other material providing a surface for deposition of a composition for IR-MALDI, for example, a stainless steel surface.

Processes for characterizing a biological macromolecule by IR-MALDI mass spectrometry are provided. For example, the mass of a biological macromolecule can be determined by preparing a composition for IR-MALDI containing the biological macromolecule to be analyzed and a liquid matrix, which absorbs infrared radiation; then analyzing the biological macromolecule in the composition by IR-MALDI mass spectrometry, thereby allowing a determination of the mass of the biological macromolecule.

A process as disclosed herein also can be used for detecting a target biological macromolecule by preparing a composition for IR-MALDI containing the target biological macromolecule and a liquid matrix, which absorbs infrared radiation, and performing IR-MALDI mass spectrometry on the composition to identify the target biological macromolecule in the composition, thereby detecting the target biological macromolecule. If desired, the target biological macromolecule can be present in or obtained from a biological sample. Accordingly, a process for identifying the presence of a target biological macromolecule in a biological sample, is provided. The presence of a target nucleic acid, for example, can be identified by preparing a composition for IR-MALDI, containing a biological sample containing nucleic acid molecules (or nucleic acid molecules isolated from the biological sample) and a liquid matrix, which absorbs infrared radiation; then analyzing the composition by IR-MALDI

mass spectrometry, wherein detection of a nucleic acid molecule having a molecular mass of the target nucleic acid sequence identifies the presence of the target nucleic acid sequence in the biological sample.

Also provided is a process of using IR-MALDI mass spectrometry to identify an individual having a disease or a predisposition to a disease by detecting a characteristic of a biological macromolecule that is obtained from the individual and is associated with the disease or the predisposition. Such a process is particularly useful for identifying a genetic disease, or a disease associated with a bacterial infection, or a predisposition to such a disease, and also is useful for determining identity, heredity or compatibility.

The processes disclosed herein are suitable for analyzing one or more target biological macromolecules, particularly a large number of target biological macromolecules, for example, by depositing a plurality of compositions, each containing one or more target biological macromolecules, on a solid support, for example, a chip, in the form of an array. The disclosed processes are particularly suitable for multiplex analysis of a plurality of biological macromolecules contained in a single composition, including a liquid matrix, in which case each biological macromolecule in the plurality can be differentially mass modified to facilitate multiplex analysis. Accordingly, the processes disclosed herein are readily adaptable to high throughput assay formats.

Processes for obtaining information on a sequence of a nucleic acid molecule by determining the identity of a target polypeptide encoded by the nucleic acid molecule are provided. In practicing these methods, a target polypeptide (or mixture thereof) is prepared from a nucleic acid molecule molecule encoding the target polypeptide; the molecular mass of the target polypeptide is determined by providing a mixture of the polypeptide with a liquid matrix, or in some embodiments, with water or succinic acid, and preforming IR-MALDI. The identity of the target polypeptide is determined by comparing the molecular mass of the target polypeptide with the molecular mass of a reference polypeptide of known identity. Information, such as the presence of a mutation, on a sequence of nucleotides in the nucleic acid molecule encoding the target polypeptide can thereby be obtained.

A biological macromolecule particularly suitable for analysis by a process of IR-MALDI mass spectrometry can be a nucleic acid, a nucleic acid analog or mimic, a triple helix, a polypeptide, a polypeptide analog or mimetic, a carbohydrate, a lipid or a proteoglycan, or can be a macromolecular complex such as a protein-protein complex or a nucleoprotein complex or other complexes. For analysis by a process as disclosed herein, a target biological macromolecule can be immobilized to a substrate, particularly a solid support, which can be, for example, a bead, a flat surface, a chip, a capillary, a pin, a comb, or a wafer, and can be any of various materials, including a metal, a ceramic, a plastic, a resin, a gel, and a membrane. Immobilization can be through a reversible linkage (i.e. an ionic bond, such as biotin/streptavidin), a covalent bond, such as photocleavable bond or a thiol linkage or a hydrogen bond, and the linkage can be cleaved using, for example, a chemical process, an enzymatic process, or a physical process, including during the IR-MALDI mass spectrometric analysis procedure.

A biological macromolecule to be analyzed can be conditioned prior to IR-MALDI mass spectrometric analysis, thereby improving the ability to analyze the particular biological macromolecule by IR-MALDI mass spectrometry, for example, by improving the resolution of the mass spectrum. A target biological macromolecule can be conditioned, for example, by ion exchange, by contact with an alkylating agent or trialkylsilyl chloride, or by incorporation of at least one mass modified subunit of the biological macromolecule. If desired, the biological macromolecule can be isolated prior to conditioning or prior to IR-MALDI mass spectrometric analysis.

A process for determining the identity of each target biological macromolecule in a plurality of target biological macromolecules, which can be fragments of a biological macromolecule, can be performed, for example, by preparing a composition for IR-MALDI containing a plurality of differentially mass modified target biological macromolecules and a liquid matrix, which absorbs infrared radiation; determining the molecular mass of each differentially mass modified target biological macromolecule in the plurality by IR-MALDI mass spectrometry; and comparing the molecular mass of each differentially

mass modified target biological macromolecule in the plurality with the molecular mass of a corresponding known biological macromolecule. Where such a process is performed using a plurality of target biological macromolecules, each of which is a fragment of a larger biological macromolecule, the fragments can be prepared by contacting the biological macromolecules with at least one agent that cleaves a bond involved in the formation of the biological macromolecules, particularly a bond between monomer subunits of the biological macromolecule.

Processes for identifying one or more subunits in a biological macromolecule using IR-MALDI mass spectrometry also are provided, for example, processes for detecting a mutation in a nucleotide sequence. The identity of a target nucleotide can be identified, for example, by hybridizing a nucleic acid molecule containing the target nucleotide with a primer oligonucleotide that is complementary to the nucleic acid molecule at a site adjacent to the target nucleotide, to produce a hybridized nucleic acid molecule; contacting the hybridized nucleic acid molecule with a complete set of dideoxynucleosides or 3'-deoxynucleoside triphosphates and a DNA dependent DNA polymerase, so that only the dideoxynucleosides or 3'-deoxynucleoside triphosphate that is complementary to the target nucleotide is extended onto the primer; preparing a composition containing the extended primer and a liquid matrix, which absorbs infrared radiation; and detecting the extended primer in the composition by IR-MALDI mass spectrometry, thereby determining the identity of the target nucleotide.

A process for detecting the absence or presence of a mutation in a target nucleic acid sequence can be performed by hybridizing a nucleic acid molecule containing the target nucleic acid sequence with at least one primer, which has 3' terminal base complementarity to the target nucleic acid sequence, to produce a hybridized product; contacting the hybridized product with an appropriate polymerase enzyme and sequentially with one of the four nucleoside triphosphates, then preparing a composition containing the reaction product and a liquid matrix, which absorbs infrared radiation; and detecting the product in the composition by IR-MALDI mass spectrometry, wherein the molecular weight

of the product indicates the presence or absence of a mutation next to the 3′ end of the primer in the target nucleic acid molecule. A mutation in a nucleic acid molecule also can be detected, for example, by hybridizing the nucleic acid molecule with an oligonucleotide probe, to produce a hybridized nucleic acid, wherein a mismatch is formed at the site of a mutation; contacting the hybridized nucleic acid with a single strand specific endonuclease, then preparing a composition containing the reaction product and a liquid matrix, which absorbs infrared radiation; and analyzing the composition by IR-MALDI mass spectrometry, wherein the presence of more than one nucleic acid fragment in the composition indicates that the nucleic acid molecule contains a mutation.

A process for identifying the absence or presence of a mutation in a target nucleic acid sequence also can be performed, for example, by performing at least one hybridization on a nucleic acid molecule containing the target nucleic acid sequence with a set of ligation educts and a DNA ligase; preparing a composition containing the reaction product and a liquid matrix, which absorbs infrared radiation; and analyzing the composition by IR-MALDI mass spectrometry. Using such a process, the detection of a ligation product in the composition identifies the absence of a mutation in the target nucleic acid sequence, whereas the detection only of the set of ligation educts in the composition identifies the presence of a mutation in the target nucleic sequence. A process of detecting the presence of a ligation product, as disclosed above, also can be useful for detecting a target nucleotide or a target nucleic acid by performing at least one hybridization on a nucleic acid molecule containing the target nucleotide with a set of ligation educts and a thermostable DNA ligase; preparing a composition containing the reaction product and a liquid matrix, which absorbs infrared radiation; and identifying a ligation product in the composition by IR-MALDI mass spectrometry, thereby detecting the presence of a target nucleotide in the nucleic acid sequence.

Processes for determining a subunit sequence of a biological macromolecule also are provided. A subunit sequence of at least one species of target biological macromolecule, i, can be determined, for example, by

contacting the species of target biological macromolecule with one or more agents sufficient to cleave each bond involved in the formation of the target biological macromolecule, to produce a set of nested biological macromolecule fragments, then preparing a composition containing at least one biological macromolecule fragment of the set and a liquid matrix, which absorbs infrared radiation; and determining the molecular mass of the at least one biological macromolecule fragment by IR-MALDI mass spectrometry; and repeating these steps until the molecular mass of each biological macromolecule fragment in the set has been determined, thereby determining the subunit sequence of the species of target biological macromolecule. Such a process is particularly suitable for multiplex analysis of a plurality of i + 1 species of target biological macromolecules, wherein each species of target biological macromolecule is differentially mass modified such that a biological macromolecule fragment of each species of target biological macromolecule can be distinguished from a biological macromolecule of each different species by IR-MALDI mass spectrometry.

Processes for determining the nucleotide sequence of at least one species of nucleic acid are provided. Such a process can be performed by synthesizing complementary nucleic acids, which are complementary to the species of nucleic acid to be sequenced, starting from an oligonucleotide primer and in the presence of chain terminating nucleoside triphosphates, to produce four sets of base-specifically terminated complementary polynucleotide fragments; preparing a composition for IR-MALDI, containing the four sets of polynucleotide fragments and a liquid matrix, which absorbs infrared radiation; determining the molecular weight value of each polynucleotide fragment by IR-MALDI mass spectrometry; and determining the nucleotide sequence of the species of nucleic acid by aligning the molecular weight values according to molecular weight. Such a process is particularly suitable to multiplex analysis of a plurality of i + 1 species of nucleic acids, which can be sequenced concurrently using i + 1 primers, wherein one of the i + 1 primers is an unmodified primer or a mass modified primer and the other i primers are mass

modified primers, and wherein each of the i + 1 primers can be distinguished from the other by IR-MALDI mass spectrometry.

A sequence of a target nucleic acid also can be determined by hybridizing at least one partially single stranded target nucleic acid to one or more nucleic acid probes, each probe containing a double stranded portion, a single stranded portion, and a determinable variable sequence within the single stranded portion, to produce at least one hybridized target nucleic acid, then preparing a composition containing the hybridized target nucleic acid and a liquid matrix, which absorbs infrared radiation; and determining a sequence of the hybridized target nucleic acid by IR-MALDI mass spectrometry based on the determinable variable sequence of the probe to which the target nucleic acid hybridized. If desired, the steps of the process can be repeated a sufficient number of times to determine an entire sequence of a target nucleic acid and, where a plurality of target nucleic acids are to be sequenced, the one or more nucleic acid probes can be immobilized in an array. If desired, the hybridized target nucleic acid can be ligated to the determinable variable sequence prior to preparing the composition for IR-MALDI.

A process for determining the sequence of a target biological macromolecule also can be performed by generating at least two biological macromolecule fragments from the target biological macromolecule, then preparing a composition containing the biological macromolecule fragments and a liquid matrix, which absorbs infrared radiation; and analyzing the biological macromolecule fragments in the composition by IR-MALDI mass spectrometry, thereby determining the sequence of the target nucleic acid molecule. Such a process is particularly useful for ordering two or more portions of a biological macromolecule sequence within a larger sequence.

Also, provided are compositions for IR-MALDI that contain a liquid matrix, which absorbs infrared radiation, and a biological macromolecule. In particular, the biological macromolecule and the liquid matrix are present in a ratio of about 10^{-4} to 10^{-9} biological macromolecule to liquid matrix in the composition. Also provided are these compositions in which the biological macromolecule is present in an amount less than about 10 picomoles of

biological macromolecule, preferably about 100 attomoles to about 1 picomole of biological macromolecule. The compositions can further include an additive that facilitates detection of the nucleic acid by IR-MALDI. Supports (or substrates) on which the compositions are deposited are provided.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A to 1C show mass spectra of a synthetic DNA 70-mer. Figure 1A shows ultraviolet matrix assisted laser desorption ionization (UV-MALDI) and detection by a linear time-of-flight (TOF) instrument using delayed extraction and a 3 hydroxypicolinic acid (3HPA) matrix (sum of 20 single shot mass spectra); Figure 1B shows UV-MALDI reflectron (ref) TOF spectrum, using delayed extraction and a 3HPA matrix (sum of 25 single shot mass spectra); Figure 1C shows IR-MALDI-refTOF spectrum, using delayed extraction and a glycerol matrix, (sum of 15 single shot mass spectra).

Figures 2A to 2D show IR-MALDI refTOF mass spectra using a 2.94μm wavelength and a glycerol matrix. The spectra are as follows: Figure 2A - a synthetic DNA 21 mer (sum of 10 single shot spectra); Figure 2B - a DNA mixture containing a restriction enzyme products of a 280-mer (87 kDA), a 360-mer (112 kDa), a 920-mer (285 kDa) and a 1400-mer (433 kDa) (sum of 10 single shot spectra); Figure 2C - DNA mixture; restriction enzyme products of a 130-mer (approximately 40 kDa), a 640-mer (198 kDa) and a 2180-mer (674 kDa) (sum of 20 single shot spectra); Figure 4D - an RNA 1206-mer (approximately 387 kDa) (sum of 15 single shot spectra). Ordinate scalings are intercomparable.

Figures 3A to 3C show the spectra of a 515-mer double stranded PCR DNA product. Total amounts of sample were loaded, as follows: Figure 3A - 300 fmol (single shot spectrum); Figure 3B - 3 fmol (single shot spectra); Figure 3C - 300 attomol (sum of 25 single shot spectra). Obtained using an IR-MALDI refTOF, wherein the laser emitted at a wavelength of 2.94 μm using a glycerol matrix.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

All patents, patent applications and publications cited herein are incorporated herein by reference. The meaning of certain terms and phrases used in the specification and claims are provided below. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which the subject matter belongs.

As used herein, a biological macromolecule refers to a molecule, which typically may be found in a biological source. Biological macromolecules include biopolymers, which are molecules containing monomeric subunits, which subunits can be the same or different. Macromolecules thus include molecules, such as peptides, proteins, small organics, oligonucleotides or monomeric units of the peptides, organics, nucleic acids and other macromolecules. A monomeric unit refers to one of the constituents from which the resulting molecule is built. Thus, monomeric units include, nucleotides, amino acids, and pharmacophores from which small organic molecules are synthesized.

Biopolymers are well known in the art and include, for example, nucleic acids, polypeptides, and carbohydrates, which are naturally occurring molecules. For purposes of the present disclosure, however, a biological macromolecule such as a biopolymer also can be a synthetic molecule that is based on or derived from a naturally occurring molecule or can be a macromolecular complex such as a nucleoprotein complex, protein-protein complex, or the like. When such molecule is a biopolymer, it contains at least one molecule containing monomeric subunits in association with a second molecule, which may or may not comprise monomeric subunits. Thus, a biopolymer can be, for example, a nucleic acid sequence containing a bond other than a phosphodiester bond between two or more nucleotides; or a polypeptide containing one or more mass modified amino acids; or a DNA

binding protein in association with a nucleic acid sequence containing the DNA binding protein recognition site or a variant thereof. The monomeric subunits of a biopolymer can be, for example, the four nucleotides that generally comprise DNA, or the twenty amino acids that generally comprise a polypeptide, or the various sugars that comprise carbohydrates, or derivatives, analogs or mimetics of such naturally occurring monomer subunits. Other biological macromolecules include lipids, glycopolypeptides, phoshpopolypeptides, peptidoglycans, oligonucleotides, polysaccharides, peptidomimetics, peptide analgos, nucleic acid analogs and other nucleic acid structures including triple helices.

As used herein, large biological macromolecules with reference to proteins refer to proteins that are approximately larger than bovine serum albumim (i.e., greater than about 65 kD).

As used herein, analyze means to identify or detect a target molecule in a sample or determination of physical or determining identifying structural characteristics, such as the presence or absence of a mutation or mass of the nucleoide, or any method in which a property of a biological macromolecule is assessed using IR MALDI.

As used herein, the term "biological sample" refers to any material obtained from a living source, for example, an animal such as a human or other mammal, a plant, a bacterium, a fungus, a protist or a virus. The biological sample can be in any form, including a solid material such as a tissue, cells, a cell pellet; a biological fluid such as urine, blood, saliva, amniotic fluid, exudate from a region of infection or inflammation; a mouth wash containing buccal cells; a cell extract, or a biopsy sample.

As used herein, the term "polymorphism" refers to the coexistence, in a population, of more than one form of an allele. A polymorphism can occur in a region of a chromosome not associated with a gene or can occur, for example, as an allelic variant or a portion thereof of a gene. A portion of a gene that exists in at least two different forms, for example, two different nucleotide sequences, is referred to as a "polymorphic region of a gene." A polymorphic region of a gene can be localized to a single nucleotide, the identity of which

differs in different alleles, or can be several nucleotides long. Of particular interest herein, polymorphisms, referred to as single nucleotide polymorphism (SNPs) that arise by virtue of change in single nucleotide base.

As used herein, the term "liquid dispensing system" means a device that can transfer a predetermined amount of liquid to a target site. The amount of liquid dispensed and the rate at which the liquid dispensing system dispenses the liquid to a target site, which can contain a reaction mixture, can be adjusted manually or automatically, thereby allowing a predetermined volume of the liquid to be maintained at the target site. Preferred dispensing systems are designed to dispense nano-liter volumes (i.e., volumes between about 1 and 100 nanoliters) of material. Such systems are known (see, *e.g.*,published International PCT application No. WO 98/20200, which is based on allowed U.S. application Serial No. 08/787,639 as well as U.S. application Serial No. 08/786,988).

As used herein, the term "liquid" is used to mean a non-solid, non-gaseous material, at room temperature and 1 atm. pressure, which can contain one or more solid or gaseous materials dissolved or suspended or otherwise mixed therein.

As used herein, the term "target site" refers to a specific locus on a solid support that can contain a liquid.

A solid support contains one or more target sites, which can be arranged randomly or in ordered array or other pattern. In particular, a target site restricts growth of a liquid to the "z" direction of an xyz coordinate. Thus, a target site can be, for example, a well or pit, a pin or bead, or a physical barrier that is positioned on a surface of the solid support, or combinations thereof such as a beads on a chip, chips in wells, or the like. A target site can be physically placed onto the support, can be etched on a surface of the support, can be a "tower" that remains following etching around a locus, or can be defined by physico-chemical parameters such as relative hydrophilicity, hydrophobicity, or any other surface chemistry that allows a liquid to grow primarily in the z direction. A solid support can have a single target site, or can contain a number of target sites, which can be the same or different, and where

the solid support contains more than one target site, the target sites can be arranged in any pattern, including, for example, an array, in which the location of each target site is defined.

As used herein, the term "target biological macromolecule" refers to any biological macromolecule of interest, including a fragment of a biological macromolecule, that is to be analyzed by IR-MALDI mass spectrometry. For example, a target biological macromolecule can be a nucleic acid such as a gene or an mRNA, or a relevant portion of a nucleic acid such as a restriction fragment or deletion fragment of the nucleic acid. A target nucleic acid can be a polymorphic region of a chromosomal nucleic acid, for example, a gene, or a region of a gene potentially having a mutation. Target nucleic acids include, but are not limited to, nucleotide sequence motifs or patterns specific to a particular disease and causative thereof, and to nucleotide sequences specific as a marker of a disease but not necessarily causative of the disease or condition. A target nucleic acid also can be a nucleotide sequence that is of interest for research purposes, but that may not have a direct connection to a disease or that may be associated with a disease or condition, although not yet proven so.

A target biological macromolecule also can be a polypeptide, or a relevant portion thereof, that is subjected to IR-MALDI mass spectrometry, for example, for identifying the presence of a polymorphism or a mutation. A target polypeptide can be encoded by a nucleotide sequence encoding a protein, which can be associated with a specific disease or condition, or a portion of a protein, or can be encoded by a nucleotide sequence that normally does not encode a translated polypeptide. A target polypeptide also can be encoded, for example, from a sequence of dinucleotide repeats or trinucleotide repeats or the like, which can be present in chromosomal nucleic acid, for example, a coding or a non-coding region of a gene, for example, in the telomeric region of a chromosome. A target polypeptide can be obtained from a naturally occurring protein or can be prepared from a nucleic acid by an *in vitro* method.

The identity of a target biological macromolecule can be determined by comparison of the molecular mass or sequence with that of a corresponding known biological macromolecule.

As used herein, the term "corresponding known biological macromolecule" means a biological macromolecule having a known characteristic, which can be any relevant characteristic including, for example, the mass or charge, the fragmentation pattern following treatment with a fragmenting agent, the tissue or cell type in which the biological macromolecule normally is found in nature, or the like. A corresponding known biological macromolecule generally is used as a control for comparison to a second biological macromolecule, particularly a target biological macromolecule. By comparing the spectra of a target biological macromolecule with a corresponding known biological macromolecule, information about the target biological macromolecule can be obtained.

As used herein, a corresponding known biological macromolecule can have substantially the same subunit sequence as the target biological macromolecule, or can be substantially different. For example, where a target polypeptide is an allelic variant that differs from a corresponding known polypeptide by a single amino acid difference, the amino acid sequences of the polypeptides will be the same except for the single difference. In comparison, where a mutation in a nucleic acid encoding the target polypeptide changes, for example, the reading frame of the encoding nucleic acid or introduces or deletes a STOP codon, the sequence of the target polypeptide can be substantially different from that of the corresponding known polypeptide.

With respect to a nucleic acid, a target nucleic acid can be, for example, a DNA molecule that is obtained from a subject, such as as prostate cancer patient and includes the polymorphic region that demonstrates amplification of a trinucleotide sequence associated with prostate cancer, and the corresponding known nucleic acid can be the same polymorphic region from a subject that does not have prostate cancer. Depending on the amount of amplification, the target nucleic acid can be substantially larger than the corresponding known nucleic acid. A target nucleic acid also can be a polymorphism or a mutated gene, which can alter the phenotype of a subject as compared to a subject not having the polymorphism or the mutated gene, and a corresponding known

nucleic acid can be the nucleotide sequence of an allele that is present in the majority of subjects in a relevant population.

A target biological macromolecule can be a fragment of a larger biological macromolecule and can be produced by contacting the larger biological macromolecule with an appropriate fragmenting agent.

As used herein, the term "fragmenting agent" means a physical, chemical or biochemical agent that, upon contacting a biological macromolecule, breaks the biological macromolecule into at least two separate portions. In general, a fragmenting agent is specific for a particular type of biological macromolecule, for example, a peptidase, which cleaves a polypeptide; a nuclease, which cleaves a nucleic acid molecule; or a glycosidase, which cleaves a carbohydrate. Non-specific fragmenting agents also are well known and include, for example, physical agents such ionizing radiation or sonication. Contacting a biological macromolecule with a fragmenting agent produces fragments of the biological macromolecule.

As used herein, the term "fragment," when used with reference to a biological macromolecule, means a portion of the biological macromolecule that has a lower molecular mass than the entire biological macromolecule. A fragment of a biological macromolecule can be one or more of the subunits that comprise the biological macromolecule, or can be portions of the biological macromolecule lacking one or more subunits, including deletion fragments.

A fragment of a polypeptide, for example, generally is produced by specific chemical or enzymatic degradation of the polypeptide. Where chemical or enzymatic cleavage occurs in a sequence specific manner, the production of fragments of a polypeptide is defined by the primary amino acid sequence of the polypeptide. Fragments of a polypeptide can be produced, for example, by contacting the polypeptide, which can be immobilized to a solid support, with a chemical agent such as cyanogen bromide, which cleaves a polypeptide at methionine residues, or hydroxylamine at high pH, which can cleave an Asp-Gly peptide bond; or with a peptidase, for example, an endopeptidase such as trypsin, which cleaves a polypeptide at Lys or Arg residues, or an exopeptidase such as carboxypeptidase, which produces one or more free amino acids, which

have been released from the carboxy terminus of the polypeptide, and deletion fragments of the polypeptide that lacks the one or more amino acids.

The term "deletion fragment" refers to a fragment of a biological macromolecule that remains following sequential cleavage of a subunit from a terminus of the biological macromolecule. The term "nested set of deletion fragments" refers to a population of deletion fragments that results from sequential cleavage of subunits from a biological macromolecule. A nested set of deletion fragments generally contains at least one deletion fragment that terminates in each subunit of at least a portion of the biological macromolecule, thereby allowing sequencing of the biological macromolecule. Thus, as many as N deletion fragments can be produced from a biological macromolecule, where "N" is the number of subunits in the biological macromolecule, although fewer than N deletion fragments can be produced. It should be recognized that a "nested set" of nucleic acid fragments also can be produced using, for example, by performing a chain-terminating polymerase reaction such as a dideoxy sequencing method.

In comparison to the production of deletion fragments using a fragmenting agent that cleaves a biological macromolecule from a terminus, treatment of a biological macromolecule with a fragmenting agent that recognizes specific sites in the biological macromolecule results in the production in M+1 fragments of the biological macromolecule, where "M" is the number of specific cleavage sites in the biological macromolecule. For example, treatment of a polypeptide having four internal and interspersed methionine residues with cyanogen bromide results in the production in five fragments of the polypeptide.

Fragments of nucleic acids, carbohydrates, or other biological macromolecules also can be produced. For example, exonucleases, including DNAses and RNAses, and endonucleases, including restriction endonucleases, can be used to produce fragments of a nucleic acid molecule (see Sambrook et al., *Molecular Cloning: A laboratory manual* (Cold Spring Harbor Laboratory Press 1989), listing nucleic acid fragmenting agents). The choice of a nuclease to produce nucleic acid fragments will depend on the process being performed

and the characteristics of the nucleic acid molecule, for example, whether it is DNA or RNA and whether, if DNA, it contains recognition sites, if necessary, for action by the nuclease. Similarly, fragments of carbohydrates can be produced using enzymes such as exoglycosidases or endoglycosidases, for example, amylases, which can produce fragments of carbohydrates containing α-1,4-glycosydic bonds (see U.S. Patent No. 5,821,063).

A nested set of deletion fragments of a target biological macromolecule can be produced using an agent that cleaves the biological macromolecule from a terminus.

As used herein, the term "agent that cleaves a biological macromolecule unilaterally from a terminus" refers to a physical, chemical or biological agent for sequentially removing subunits from one end of a biological macromolecule. A biological agent that cleaves a biological macromolecule unilaterally from a terminus is exemplified by an exopeptidase such as carboxypeptidase Y, which sequentially cleaves amino acids from the carboxyl terminus of a polypeptide (see U.S. Patent No. 5,792,664; International Publ. WO 96/36732), or by an exonuclease such as exonuclease III, which sequentially cleaves nucleotides from the 3'-hydroxyl terminus of a double stranded DNA (see International Publ. WO 94/21822). A physical agent is exemplified by a light source, for example, a laser, which can cleave a terminal subunit from a biological macromolecule, particularly where the subunit is bound to the biological macromolecule through a photolabile bond. A chemical agent is exemplified by phenylisothiocyanate (Edman's reagent), which, in the presence of an acid, cleaves an amino terminal amino acid from a polypeptide.

As used herein, the residues of naturally occurring α-amino acids are the residues of those 20 α-amino acids found in nature that are incorporated into protein by the specific recognition of the charged tRNA molecule with its cognate mRNA codon in humans.

As used herein, non-naturally occurring amino acids refer to amino acids that are not genetically encoded. Preferred such non-naturally occurring amino acids herein include those with unsaturated side chains.

As used herein, the term "polypeptide" means at least two amino acids, or amino acid derivatives, which can be mass modified amino acids or non-naturally-occurring amino acids, that are linked by a peptide bond, which can be a modified peptide bond. Exemplary polypeptides include, but are not limited to, native proteins, gene products, protein conjugates, mutant or polymorphic polypeptides, post-translationally modified proteins, genetically engineered gene products including products of chemical synthesis, *in vitro* translation, cell-based expression systems, including fast evolution systems involving vector shuffling, random or directed mutagenesis and peptide sequence randomization, oligopeptides, antibodies, enzymes, receptors, regulatory proteins, nucleic acid-binding proteins, hormones, or protein products of a display method such as phage or bacterial display methods.

A polypeptide can be translated from a nucleotide sequence that is at least a portion of a coding sequence, or from a nucleotide sequence that is not naturally translated due, for example, to its being in a reading frame other than the coding frame or to its being an intron sequence, a 3′ or 5′ untranslated sequence, or a regulatory sequence such as a promoter. A polypeptide also can be chemically synthesized and can be modified by chemical or enzymatic methods following translation or chemical synthesis. The terms "protein," "polypeptide" and "peptide" can be used interchangeably herein when referring to a translated nucleic acid, for example, a gene product, although "peptides" generally are smaller than "polypeptides" and "proteins" often can have post-translational modifications.

As used herein, the term "nucleic acid" refers to a polynucleotide containing at least two covalently linked nucleotide or nucleotide analog subunits. A nucleic acid can be a deoxyribonucleic acid (DNA), a ribonucleic acid (RNA), or an analog of DNA or RNA, such as PNA, and can contain, for example, one or more nucleotide analogs or a covalent linkage (backbone) other than a phosphodiester bond, for example, a thioester bond, a phosphotriester bond, or a peptide bond (peptide nucleic acid; PNA; see, for example, Tam et al., Nucl. Acids Res. 22:977-986 (1994); Ecker and Crooke, BioTechnology 13:351360 (1995)); triple helices are also contemplated. The nucleic acid can

be single-stranded, double-stranded, or a mixture thereof. For purposes herein, unless specified otherwise, the nucleic acid is double-stranded or it is apparent from the context. Nucleotide analogs are commercially available and methods of preparing polynucleotides containing such nucleotide analogs are well known (Lin et al., Nucl. Acids Res. 22:5220-5234 (1994); Jellinek et al., Biochemistry 34:11363-11372 (1995); Pagratis et al., Nature Biotechnol. 15:68-73 (1997)).

A nucleic acid can be single stranded or double stranded, including, for example, a DNA-RNA hybrid. A nucleic acid also can be a portion of a longer nucleic acid molecule, for example, a portion of a gene containing a polymorphic region. The molecular structure of a nucleic acid, for example, a gene or a portion thereof, is defined by its nucleotide content, including deletions, substitutions or additions of one or more nucleotides; the nucleotide sequence; the state of methylation; or any other modification of the nucleotide sequence. Although a nucleic acid contains two or more nucleotides or nucleotide analogs linked by a covalent bond, including single stranded or double stranded molecules, it should be recognized that a "fragment" of a nucleic acid, which can be produced as discussed above, can be as small as a single nucleotide. The terms "polynucleotide" and "oligonucleotide" also are used herein to mean two or more nucleotides or nucleotide analogs linked by a covalent bond, although oligonucleotides such as PCR primers generally are less than about fifty to one hundred nucleotides in length.

As used herein, the phrase "determining the identity of a target biological macromolecule" refers to determining at least one characteristic of the biological macromolecule, which can be a nucleic acid, polypeptide or other biological macromolecule. Determining the identity of a biological macromolecule can include, for example, determining the molecular mass or charge of the biological macromolecule; or determining the identity of at least one subunit, or of a subunit sequence of the biological macromolecule; or determining a particular pattern of fragments of the biological macromolecule. For example, where the biological macromolecule is a nucleic acid, determining the identity of the target nucleic acid can include determining at least one nucleotide of the target nucleic acid, or determining the number of nucleotide

repeats present in a sequence of tandem nucleotide repeats. Similarly, where the target biological macromolecule is a polypeptide, determining the identity of the target polypeptide can include determining at least one amino acid, or a particular pattern of peptide fragments of the target polypeptide, for example, following treatment of the polypeptide with an endopeptidase. Determining the identity of a target biological macromolecule is performed by subjecting the target biological macromolecule, if necessary, to a particular reaction, as appropriate; preparing a composition containing target biological macromolecule or reaction product thereof and a liquid matrix, which absorbs IR radiation; and analyzing the target biological macromolecule or reaction product thereof by IR-MALDI mass spectrometry.

The terms "infrared radiation" and "infrared wavelength" refer to electromagnetic wavelengths that are longer than those of red light in the visible spectrum and shorter than radar waves, generally wavelengths within the range of about 760 nm to about 50 μm. An appropriate infrared wavelength can be generated using a laser, as disclosed herein.

As used herein, the term "liquid matrix" means a material that has a sufficient absorption at the wavelength of the laser to be used in performing desorption and ionization (i.e. an IR emitting laser) and that is a liquid at room temperature (about 20°C, 1 atm). The contemplated liquids are those that can form vitreous solids or glasses in the solid state as opposed to a crystalline structure, such as that which forms when a matrix such as picolinic acid or 3HPA is dried. Vitreous solids and glasses do not form solid crystalline heterogenous structures, but rather retain properties of liquids that derive from their lack of ordered structure. In addition, such liquid matrices form a homogenous layer when applied to the surface of a substrate or support. Thus, for purposes herein, liquid matrices are relatively non-volatile materials that are biocompatible, particularly compatible with nucleic acids and/or proteins, and include, but are not limited to, alcohols, including glycols and polyols, such as glycerol, sugars, such as sucrose, mannose, galactose, and other sugars as well as polymeric sugars, ethylene glycol, propylene glycol, trimethylolpropane, pentaerythritol, dextrose, methylglycoside or sorbitol. sucrose, mannose and

other such materials that in the solid state can form glasses rather than crystalline structures. Also included is "glassy" water, which state occurs under conditions in which very small volumes, i.e., submicroliters, particular nanoliters or less, are dispensed. Other liquid matrices include, but are not limited to triethanolamine, lactic acid, 3-nitrobenzylalcohol, diethanolamine, DMSO, nitropheynloctylether (3-NPOE), 2,2'dithiodiethanol, tetraethyleneglycol, dithiotrietol/erythritol (DTT/DTE), 2,3-dihydroxy-propyl-benzyl ether, *α*-tocopherol, and thioglycerol. Other suitable "liquid" matrices are set forth below.

For absorption purposes, the liquid matrix can contain at least one chromophore or functional group that strongly absorbs infrared radiation. Examples of appropriate functional groups include nitro, sulfonyl, sulfonic acid, sulfonamide, nitrile or cyanide, carbonyl, aldehyde, carboxylic acid, amide, ester, anhydride, ketone, amine, hydroxyl, aromatic rings, dienes and other conjugated systems. A liquid matrix, which absorbs IR radiation, including a composition containing a biological macromolecule to be analyzed by IR-MALDI and a liquid matrix, can contain an additive that facilitates IR-MALDI analysis of the biological macromolecule.

As used herein, appropriate viscosity, refers to the viscosity for dispensing glass-type liquid matrices and means that it can be dispensed as a small volume and evenly distribute over a small surface area in an thin layer.

As used herein, the term "additive" means a material that facilitates IR-MALDI analysis of a biological macromolecule. For example, an additive can facilitate solubility of the biological macromolecule in a composition containing a liquid matrix. An additive also can be a compound or compounds that have a high extinction coefficient (E) at the laser wavelength used for desorption and ionization, for example, dinitrobenzenes or polyenes. Additives also include compounds that alter the ionic strength of the matrix/sample mixture or the matrix. Exemplary salt additives include, but are not limited to, ammonium

salts and salts of amines. Exemplary salt additives for this purpose include NH_4-acetate and Tris-HCl.

Where the biological macromolecule to be analyzed by IR-MALDI is a nucleic acid, for example, an additive can be a compound that acidifies the liquid matrix, thereby inducing dissociation of double stranded nucleic acids or denaturing a secondary structure of a nucleic acid such as tRNA or other single stranded nucleic acid. An additive also can minimize salt formation between the matrix and the biological macromolecule and can be, for example, a material that conditions the biological macromolecule. When it is desirable to analyze or detect a double-stranded nucleic acid by IR-MALDI, the additive can be a substance that stablizes the double-stranded molecule or reduces denaturation of the double-stranded nucleic acid, but that is generally compatible with mass spectrometric analysis. Such additives include, but are not limited to, salts. Preferred salt additives include ammonium salts and salts of amines. Exemplary salt additives for this purpose include NH_4-acetate and Tris-HCl.

The matrix can be treated by further purification to remove other organic contaminants, including harmful derivatives and other by-products of the production process.

A biological macromolecule or fragment thereof, particularly a target biological macromolecule, can be conditioned prior to IR-MALDI mass spectrometry.

As used herein, the term "conditioned" or "conditioning," when used in reference to a biological macromolecule, means that the biological macromolecule is modified so as to decrease the amount of IR radiation required to ionize or volatilize the biological macromolecule, to minimize the likelihood of undesirable fragmentation of the biological macromolecule, or to increase the resolution of a mass spectrum of the biological macromolecule or fragments thereof. Resolution of a mass spectrum of a target biological macromolecule or fragment thereof can be increased by conditioning the biological macromolecule prior to performing IR-MALDI mass spectrometry. Conditioning can be performed at any stage prior to IR-MALDI mass spectrometry, particularly while the biological macromolecule is immobilized to a substrate. Conditioning

includes any process that achieves these results, and includes, but is not limited to, subjecting the macromolecule to ion exchange or other process that provides for a uniform charge distribution, mass modification, modification of the phosphodiester backbone of a nucleic acid, removal of negative charge from the phosphodiester backbone, cation exchange, further purification, and any other such process known to those of skill in the art to achieve conditioning.

Conditioning of a biological macromolecule will depend, in part, on the biochemical nature of the biological macromolecule. For example, a biological macromolecule can be conditioned by treatment with a cation exchange material or an anion exchange material, which reduces the charge heterogeneity of the biological macromolecule, thereby eliminating peak broadening due to heterogeneity in the number of cations (or anions) bound to the target biological macromolecule. A polypeptide, for example, can be conditioned by treatment with an alkylating agent such as alkyliodide, iodoacetamide, iodoethanol, or 2,3-epoxy-1-propanol, which prevents the formation of disulfide bonds. Such alkylating agents also can be used to condition a nucleic acid by transforming the monothiophosphodiester bonds to phosphotriester bonds. A polypeptide also can be conditioned by converting charged amino acid side chains to uncharged derivatives by contact with trialkylsilyl chlorides, which also can be used to condition a nucleic acid by transforming phosphodiester bonds to uncharged derivatives. Biological macromolecules also can be conditioned by incorporating modified subunits that are more stable than the corresponding unmodified subunits, for example, the substitution of N7- or N9-deazapurine nucleotides in a target nucleic acid, thereby minimizing the likelihood of fragmentation of the biological macromolecule.

The processes disclosed herein provide methods for analyzing a plurality of biological macromolecules in one or a few samplings, for example, by multiplex analysis.

As used herein, the term "multiplex" refers to simultaneously determining the identity of at least two target biological macromolecules by IR-MALDI mass spectrometry. For example, where a population of different target biological macromolecules are present in an array on a microchip or other substrate,

multiplexing can be used to determine the identity of a plurality of target biological macromolecules. Multiplexing can be performed, for example, by differentially mass modifying each different biological macromolecule of interest, then using IR-MALDI mass spectrometry to determine the identity of each different biological macromolecule. Multiplex analysis provides the advantage that a plurality of target biological macromolecules can be identified in as few as a single IR-MALDI mass spectrum, as compared to having to perform a separate mass spectrometric analysis for each individual target biological macromolecule.

"Multiplexing" can be achieved by several different methodologies. For example, several mutations can be simultaneously detected on one target sequence by employing corresponding detector (probe) molecules (e.g. oligonucleotides or oligonucleotide mimetics). The molecular weight differences between the detector oligonucleotides D1, D2 and D3 must be large enough so that simultaneous detection (multiplexing) is possible. This can be achieved either by the sequence itself (composition or length) or by the introduction of mass-modifying functionalities into the detector oligonucleotide. Mass modifying moieties can be attached, for instance, to either the 5'-end of the oligonucleotide, to the nucleobase (or bases), to the phosphate backbone, and to the 2'-position of the nucleoside (nucleosides) or/and to the terminal 3'-position. Examples of mass modifying moieties include, for example, a halogen, an azido, or of the type, XR, wherein X is a linking group and R is a mass-modifying functionality. The mass-modifying functionality can thus be used to introduce defined mass increments into the oligonucleotide molecule.

The mass-modifying moiety, M, can be attached either to the nucleobase, in case of, for example, c^7-deazanucleosides also to C-7, to the triphosphate group at the alpha phosphate, or to the 2'-position of the sugar ring of the nucleoside triphosphate. Furthermore, the mass-modifying functionality can be added so as to affect chain termination, such as by attaching it to the 3'-position of the sugar ring in the nucleoside triphosphate. As another examplary embodiment, various mass-modifying functionalities, R, other than oligo/polyethylene glycols, can be selected and attached via appropriate linking chemistries, X. A simple mass-modification can be achieved

by substituting H for halogens like F, Cl, Br and/or I, or pseudohalogens such as SCN, NCS, or by using different alkyl, aryl or aralkyl moieties such as methyl, ethyl, propyl, isopropyl, t-butyl, hexyl, phenyl, substituted phenyl, benzyl, or functional groups such as CH_2F, CHF_2, CF_3, $Si(CH_3)_3$, $Si(CH_3)_2(C_2H_5)$, $Si(CH_3)(C_2H_5)_2$, $Si(C_2H_5)_3$. Yet another mass-modification can be obtained by attaching homo- or heteropeptides through the nucleic acid molecule (e.g. detector (D)) or nucleoside triphosphates. One example useful in generating mass-modified species with a mass increment of 57 is the attachment of oligoglycines, e.g. mass-modifications of 74 (r = 1, m = 0), 131 (r = 1, m = 2), 188 (r = 1, m = 3), 245 (r = 1, m = 4) are achieved. Simple oligoamides also can be used, e.g., mass-modifications of 74 (r = 1, m = 0), 88 (r = 2, m = 0), 102 (r = 3, m = 0), 116 (r = 4, m = 0), etc. are obtainable. . The mass modifications serve, not only to aid in multiplexing, but to enhance or aid in resolving mass spectrometry of fragments (i.e., mass modification aids in "conditioning" the nucleic acids for analyis. Other chemistries can be used in the mass-modified compounds, as for example, those described in *Oligonucleotides and Analogues, A Practical Approach, F. Eckstein*, editor, IRL Press, Oxford, 1991 and are known to those of skill in the art of mass spectrometry.

As used herein, the term "plurality," when used in reference to biological macromolecules, means two or more biological macromolecules, each of which has a different subunit sequence. The difference in sequences can be due to a naturally occurring variation among the sequences, for example, to an allelic variation in a nucleotide or an encoded amino acid, or can be due to the introduction of particular modifications into various sequences, for example, the differential incorporation of mass modified nucleotides or amino acids into each nucleic acid or polypeptide, respectively, in the plurality.

The processes as disclosed herein can be performed using an isolated biological macromolecule.

As used herein, the term "isolated" means that a biological macromolecule is substantially separated from macromolecules normally associated with the biological macromolecule in its natural state. An isolated

nucleic acid molecule, for example, is substantially separated from the cellular material normally associated with it in a cell or, as relevant, can be substantially separated from bacterial or viral material; or from culture medium where produced by recombinant DNA techniques; or from chemical precursors or other chemicals where the nucleic acid is chemically synthesized. In general, an isolated nucleic acid molecule, which can be a fragment of a larger nucleic acid, is at least about 50% enriched with respect to its natural state, and generally is about 70% to about 80% enriched, particularly about 90% or 95% or more. Preferably, an isolated nucleic acid constitutes at least about 50% of a sample containing the nucleic acid, and can be at least about 70% or 80% of the material in a sample, particularly at least about 90% to 95% or greater of the sample.

Similarly, an isolated polypeptide can be identified based on its being enriched with respect to materials it naturally is associated with or its constituting a fraction of a sample containing the polypeptide to the same degree as defined above, i.e., enriched at least about 50% with respect to its natural state or constituting at least about 50% of a sample containing the polypeptide. An isolated polypeptide, for example, can be purified from a cell that normally expresses the polypeptide or can produced using recombinant DNA methodology, and can be a fragment of a larger polypeptide.

A biological macromolecule can be isolated using a reagent that interacts specifically with the biological macromolecule or with a tag attached to the biological macromolecule. For example, a target polypeptide can be isolated using a reagent that interacts specifically with the target polypeptide, with a peptide tag (i.e. peptide that can serve to specifically bind to a reagent, such as a column) fused to the target polypeptide, or with a peptide tag conjugated to the target polypeptide.

As used herein, the term "reagent" means a ligand or a ligand binding molecule that interacts specifically with a particular ligand binding molecule or ligand, respectively. The term "tag peptide" or "peptide tag" is not to be confusedwith a mass tag, and is used herein to mean a peptide, for which a

reagent is available. The term "tag" refers more generally to any molecule, for which a reagent is available and, therefore, includes a tag peptide.

As used herein, reagent can be an antibody that interacts specifically with an epitope of a target biological macromolecule, for example, a polypeptide, or with an epitope of a tag attached to the target biological macromolecule. For example, a reagent can be an anti-myc epitope antibody, which can interact specifically with a myc epitope fused to a target polypeptide. A reagent also can be, for example, a metal ion such as nickel ion or cobalt ion, which interacts specifically with a polyhistidine tag peptide; or zinc, copper or, for example, a zinc finger domain, which interacts specifically with a polyarginine or polylysine tag peptide; or a molecule such as avidin, streptavidin or a derivative thereof, which interacts specifically with a tag such as biotin or a derivative thereof (see International Publ. WO 97/43617, which describes, for example, methods for dissociating biotin compounds, including biotin and biotin analogs conjugated (biotinylated) to a polypeptide, from biotin binding compounds, including avidin and streptavidin, using amines, particularly ammonia).

A tag such as biotin also can be incorporated into a target nucleic acid, thereby allowing isolation of the target nucleic acid using a reagent such as avidin or streptavidin. In addition, a target nucleic acid can be isolated by hybridization to reagent containing a complementary nucleic acid sequence, which can be immobilized to a solid support such as beads, for example, magnetic beads, if desired.

The term "interacts specifically," when used in reference to a reagent and a target biological macromolecule sequence or a tag to which the reagent binds, indicates that binding occurs with relatively high affinity. As such, a reagent has an affinity of at least about $1 \times 10^6\ M^{-1}$, generally, at least about $1 \times 10^7\ M^{-1}$, and, in particular, at least about $1 \times 10^8\ M^{-1}$, for the particular biological macromolecule sequence or tag. A reagent the interacts specifically, for example, with a particular tag peptide primarily binds the tag peptide, regardless of whether other unrelated molecules are present and, therefore, is useful for isolating the tag peptide, including a target polypeptide fused to the

tag peptide, from a sample containing the target polypeptide, for example, from an *in vitro* translation reaction. Similarly, a reagent complementary nucleic acid sequence that interacts specifically with a target nucleic acid selectively binds the target nucleic acid, but not unrelated nucleic acid molecules.

A hybridizing nucleic acid sequence, which generally is an oligonucleotide, is at least nine nucleotides in length, such sequences being particularly useful as primers for the polymerase chain reaction (PCR), and can be at least fourteen nucleotides in length or, if desired, at least seventeen nucleotides in length, such nucleotide sequences being particularly useful as hybridization probes, as well as for PCR. It should be recognized that the conditions required for specific hybridization of an oligonucleotide, for example, a PCR primer, with a nucleic acid sequence, for example, a target nucleic acid, depends, in part, on the degree of complementarity shared between the sequences, the GC content of the hybridizing molecules, and the length of the antisense nucleic acid sequence, and that conditions suitable for obtaining specific hybridization can be calculated based on readily available formulas or can be determined empirically (Sambrook et al., *Molecular Cloning: A laboratory manual* (Cold Spring Harbor Laboratory Press 1989); Ausubel et al., Current Protocols in Molecular Biology (Green Publ., NY 1989)).

It can be advantageous in performing a disclosed process to immobilize a biological macromolecule, for example, a target nucleic acid or a target polypeptide, on a a substrate, particularly a solid support, such as a bead, microchip, glass or plastic capillary, or any surface, particularly a flat surface, which can contain a structure such as wells, pins or the means by which the target macromolecule is constrained at a site. A biological macromolecule can be conjugated to a solid support by various means, including, for example, by a streptavidin or avidin to biotin interaction; a hydrophobic interaction; by a magnetic interaction using, for example, functionalized magnetic beads such as DYNABEADS, which are streptavidin coated magnetic beads (Dynal Inc.; Great Neck NY); by a polar interaction such as a "wetting" association between two polar surfaces or between oligo/polyethylene glycol; by the formation of a covalent bond such as an amide bond, a disulfide bond, a thioether bond, or the

like; through a crosslinking agent; and through an acid-labile or photocleavable linker (see, for example, Hermanson, *Bioconjugate Techniques* (Academic Press 1996)). In addition, a tag can be conjugated to biological macromolecule of interest, particularly to a target biological macromolecule.

As used herein, the term "conjugated" or "immobilized" refers to an attachment, which can be a covalent attachment or a noncovalent attachment, that is stable under defined conditions. As disclosed herein, a biological macromolecule can be immobilized to a substrate, or a first substrate can be conjugated to second substrate. Immobilization of a biological macromolecule to a substrate can be direct or can be indirect through a linker, and can reversible or irreversible. A reversible immobilization can be reversed either by cleaving the attachment, for example, using light to cleave a photocleavable bond, or by subjecting the attachment to conditions that reverse the bond, for example, reducing conditions, which reverse a disulfide linkage.

As used herein, the term "substrate" or "solid support" means a flat surface or a surface with structures, to which a functional group, including a biological macromolecule containing a reactive group, can be conjugated. The term "surface with structures" means a substrate that contains, for example, wells, pins or the like, to which a functional group, including a biological macromolecule containing a reactive group, can be attached. Numerous examples of solid supports (substrates) are disclosed herein or otherwise known in the art.

A process as disclosed herein can be used to identify a subject that has or is predisposed to a disease or condition. As used herein, the term "disease" has its commonly understood meaning of a pathologic state in a subject. For purposes of the present disclosure, a disease can be due, for example, to a genetic mutation, a chromosomal defect or an infectious organism. The term "condition," which is to be distinguished from conditioning of a biological macromolecule, is used herein to mean any state of a subject, including, for example, a pathologic state or a state that determines, in part, how the subject will respond to a stimulus. The condition of a subject can be determined, in part, by determining a characteristic of the subject's genotype, which can

provide an indication as to how the subject will respond, for example, to a graft or to treatment with a particular medicament; or by detecting a particular biological macromolecule in a biological sample obtained from the subject, for example, expression of a carbohydrate associated with a particular disease. Accordingly, reference to a subject being predisposed to a condition can indicate, for example, that the subject has a genotype indicating that the subject will not respond favorably to a particular medicament or that the subject will reject a particular graft.

Reference herein to an allele or an allelic variant being "associated" with a disease or condition means that the particular genotype is characteristic, at least in part, of the genotype exhibited by a population of subjects that have or are predisposed to the disease or condition. For example, an allelic variant such as a mutation in the BRCA1 gene is associated with breast cancer, and an allelic variant such as a higher than normal number of trinucleotide repeats in a particular gene is associated with prostate cancer. The skilled artisan will recognize that an association of an allelic variant with a disease or condition can be identified using well known statistical methods for sampling and analysis of a population.

As used herein, compositions include mixtures of materials and as well as solutions.

Except as otherwise disclosed, the practice of the processes described herein employs conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art and described, for example, in *DNA Cloning*, Volumes I and II (D.N. Glover, ed., 1985); *Oligonucleotide Synthesis* (M.J. Gait, ed., 1984); Mullis et al., U.S. Patent No: 4,683,194; *Nucleic Acid Hybridization* (Hames and Higgins, eds., 1984); *Transcription and Translation* (Hames and Higgins eds., 1984); *Culture of Animal Cells* (R.I. Freshney; Alan R. Liss, Inc., 1987); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); *Gene Transfer Vectors For Mammalian Cells* (Miller and Calos, eds.; Cold Spring Harbor Laboratory 1987); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al., eds., Academic Press, NY),

Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds.; Academic Press London, 1987); *Handbook Of Experimental Immunology*, Volumes I to IV (Weir and Blackwell, eds., 1986); *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory press, Cold Spring Harbor NY, 1986).

PROCESSES AND COMPOSITIONS FOR USE WITH IR MALDI

The processes and compositions disclosed herein allow the detection, identification or characterization of biological macromolecules, including nucleic acids, polypeptides, and carbohydrates, as well as macromolecular complexes such as protein complexes and nucleoprotein complexes, by infrared (IR) matrix assisted laser desorption/ionization (MALDI) mass spectrometry. A composition for IR-MALDI is provided, the composition being a composition containing at least a biological macromolecule to be analyzed by IR-MALDI mass spectrometry and a liquid matrix, which absorbs IR radiation. Such a composition, which can be deposited on a substrate, is useful for determining a characteristic of a biological macromolecule by IR-MALDI mass spectrometry.

Processes for analyzing a target biological macromolecule using IR-MALDI mass spectrometry also are provided, including, for example, processes for detecting a target biological macromolecule in a sample, particularly a biological sample; processes for determining the identity of a biological macromolecule such as the presence of a mutation or other genetic change in a nucleic acid or of an amino acid change in a polypeptide encoded by a nucleic acid having a genetic change; and processes for determining a sequence of a biological macromolecule. The processes disclosed herein allow the analysis by IR-MALDI mass spectrometry of one or more target biological macromolecules, either in separate, but related processes such as a high throughput process, where the biological macromolecules can be analyzed serially, or can be arranged in an array on a silicon wafer, for example, and analyzed in parallel; or in a single process using a multiplex format, where each biological macromolecule in a plurality is differentially identifiable, for example, due to differential mass modification of the biological macromolecules.

The disclosed processes and compositions are based, in part, on the finding that high resolution mass spectra of large nucleic acid molecules (DNA

and RNA) can be obtained by desorbing and ionizing the nucleic acids in a liquid matrix using a laser that emits in the infrared electromagnetic wavelength. Accordingly, a process is provided for performing IR-MALDI mass spectrometry, containing mixing a nucleic acid composition with a liquid matrix to form a matrix/nucleic acid composition and depositing the composition onto a substrate to form a homogeneous, thin layer of matrix/nucleic acid composition. The nucleic acid containing substrate then can be illuminated with IR radiation of an appropriate wavelength to be absorbed by the matrix, so that the nucleic acid is desorbed and ionized, thereby emitting ion particles that can be extracted (separated) and analyzed by a mass analyzer to determine the mass of the nucleic acid. A process for analyzing a nucleic acid by mass spectrometry can be performed by depositing a composition containing the nucleic acid and a liquid matrix on a substrate, to form a homogeneous, thin layer of a nucleic acid/liquid matrix composition; illuminating the substrate containing the deposited composition with an infrared laser, so that the nucleic acid is desorbed and ionized; and mass separating and detecting the ionized nucleic acid using an appropriate mass separation and analysis format.

Processes are provided for analyzing a target biological macromolecule, particularly a target nucleic acid, by preparing a composition containing the target biological macromolecule and a liquid matrix, which absorbs IR radiation, and analyzing the target biological macromolecule in the composition by IR-MALDI mass spectroscopy. The various processes disclosed herein allow a determination of the molecular mass of a target biological macromolecule, the detection or identification of a target biological macromolecule, which can be present in a biological sample, or the determination of a subunit sequence of a target biological macromolecule. Depending on the source of the target biological macromolecule, a process as disclosed herein can be useful, for example, for determining whether an individual has a disease or a predisposition to a disease, or for determining heredity, identity or compatibility of an individual (see International Publ. WO 98/20019).

A target biological macromolecule, for example, a target nucleic acid molecule, can be obtained from a subject, particularly from a cell or tissue in the

subject or from a biological fluid, i.e., a biological sample. A target biological macromolecule can be a target nucleic acid molecule, or can be a target polypeptide, which can be obtained, for example, by *in vitro* translation of an RNA molecule encoding the target polypeptide; or by *in vitro* transcription of a nucleic acid encoding the target polypeptide, followed by translation, which can be performed *in vitro* or in a cell, where the nucleic acid to be transcribed is obtained from a subject. The processes disclosed herein provide fast and reliable methods for identifying or obtaining information about the target biological macromolecule.

Exemplary Advantages of IR-MALDI in the Detection of Target Molecules Obtained from Biological Samples

Biological samples containing a target molecule which have undergone some purification still are likely to contain extraneous contaminants (i.e., materials other than the target molecule) that are not present in a pure sample of target molecule. For example, extraneous proteins and salts may be present in partially purified preparations thereby making such preparations in reality "mixtures" as opposed to pure samples. Accordingly, mass resolution, accuracy, sensitivity and the signal-to-noise ratio become very critical parameters in mass spectrometric methods designed to detect the presence of a target molecule obtained from a biological sample. The mass spectrometric technique must be able to clearly resolve the target molecule, which may not be present in significant quantities, from the contaminant materials.

Thus, the fact that a particular mass spectrometric method may be used to measure the mass of a relatively pure biological molecule is no guarantee that it will be applicable to the detection of target molecules obtained from a biological sample. Furthermore, because of the inherent differences in the various types of mass spectrometric methods (e.g., ESI and MALDI using different lasers and/or matrices), the fact that one mass spectrometric technique may be useful in the detection of target molecules obtained from a biological sample is no guarantee that another type would also be suitable for this purpose. Additionally, the fact that a particular mass spectrometric method or set of conditions may be used to detect one particular type of target molecule,

from a biological sample does not guarantee that it can be used effectively to detect another type of target molecule from a biological sample. For example, even different sizes and types of a single class of target molecule (e.g., single-stranded vs. double-stranded DNA) from a biological sample may or may not be detected by different mass spectrometric methods and conditions, just as completely different classes of target molecules, e.g., nucleic acids vs. proteins, from a biological sample may or may not be detected by different mass spectrometric methods and conditions.

A comparison of proteins and nucleic acids reveals several differences that directly impact their amenability to analysis by mass spectrometry. For example, nucleic acids are typically more susceptible to fragmentation than proteins due to losses of nucleobases as a result of the labile N-glycosidic bond between the different bases and the deoxyribose moiety and to depurination. Spectra of nucleic acids reveal a greater tendency toward adduct formation than those of proteins. Furthermore, the relative ease of desorption/ionization appears to be greater for proteins as compared to nucleic acids since proteins tend to fold into defined structures whereas nucleic acids have less tertiary structure than proteins.

As disclosed herein, IR-MALDI mass spectrometry has been found to be effective and advantageous in methods of detection of target molecules, particularly large target molecules, obtained from biological samples. This has been due in part to the recognition of the significance of defining the optimal parameters (for example, the particular combinations of laser, wavelength, matrix, additive, pulse width, beam profile, temperature and/or fluence) that provide the level of resolution, sensitivity, signal-to-noise level, etc., required to detect a target molecule obtained from a biological sample.

For example, shorter pulse widths can be used in IR-MALDI mass spectrometric detection of target molecules, particularly employing lasers with optoelectronic switches. Typically, pulse widths less than about 90 ns, and generally about 80 ns, may be used in IR-MALDI mass spectrometric detection methods.

In addition, lower electric field strength for ion extraction can be used in IR-MALDI mass spectrometric detection of target molecules. Field strengths of about less than 1000V/mm to about 200 V/mm may typically be used in IR-MALDI mass spectrometric detection of target molecules. Furthermore, the single-shot ion signals are a factor of 3-5 times more intense than those obtained with UV-MALDI mass spectrometry, and fewer shots may be required to obtain an adequate signal-to-noise ratio.

With these improvements, the choice of laser fluence (energy per unit area on the sample) can be much less critical. Whereas in order to avoid risking substantial ion fragmentation in UV-MALDI mass spectrometry it is necessary to restrict fluence to values between H_0 and 1.5 H_0, in the disclosed IR-MALDI mass spectrometric methods for detecting target molecules, it is possible to use fluence values of up to 3 H_0 or 5 H_0, particularly when glycerol is used as a matrix.

In addition, glycerol, when used as a matrix in IR-MALDI mass spectrometry has been found to be particularly tolerant to contaminants such as salts, buffers, detergents, etc. in the sample being analyzed for the presence or absence of a target molecule. This has been surprisingly advantageous in the detection of target polypeptides, particularly large polypeptides, by IR-MALDI mass spectrometry using glycerol as a matrix because polypeptides obtained from biological samples can contain such contaminants. Such contaminants, for instance, salts, can interfere with UV-MALDI measurement of polypeptides using more traditional acidic solid state matrices. Accordingly, less purification of target molecules from biological samples is required in preparing a sample for analysis by IR-MALDI using a glycerol matrix than by UV-MALDI.

For a glycerol matrix, when used in IR-MALDI mass spectrometric methods, the molar ratio of analyte-to-matrix is much less critical than it is for crystalline matrices. Analyte-to-matrix ratios in the range of about 5×10^{-3} and 1×10^{-6} can be employed in IR-MALDI mass spectrometric detection of target molecules without substantial degradation of the ion signal. This is particularly advantageous in the analysis of biological samples when the concentration of target molecule may not be known.

With these improved conditions and other conditions and methods as described herein, clear ion signals for even large, e.g., greater than 500 kDa proteins and greater than 700 kDa nucleic acids, target molecules from biological samples are obtainable using IR-MALDI mass spectrometry. Thus, the detection of target molecules, particularly large target molecules, obtained from biological samples notoriously difficult to analyze due to the presence of mixtures, contaminants, impurities is made possible by IR-MALDI mass spectrometry and further is made amenable to automation as desired in large-scale diagnostic and screening procedures.

COMPOSITIONS FOR IR-MALDI ANALYSIS OF BIOLOGICAL MACROMOLECULES

Compositions, which are suitable for IR-MALDI, are provided herein. Such a composition referred to herein as a "composition for IR-MALDI," is a liquid mixture containing a biological macromolecule, which is to be analyzed by IR-MALDI, and a liquid matrix, which absorbs infrared radiation. A biological macromolecule suitable for analysis by IR-MALDI can be, for example, a nucleic acid, a polypeptide or a carbohydrate, or can be a macromolecular complex such as a nucleoprotein complex, protein-protein complex, a polypsaccharide, an oligosaccharide, such as dextrans and dextrins, lipids, lipopolysaccharides and other macromolecules.

A composition for IR-MALDI contains the biological macromolecule, for example, a nucleic acid, and the liquid matrix, generally in a ratio of about 10^{-4} to 10^{-9}. The composition for IR-MALDI and can contain less than about 10 picomoles of biological macromolecule to be analyzed, for example, about 100 attomol to about 1 picomole (pmol) of the biological macromolecule. A composition for IR-MALDI also can contain an additive, which facilitates detection of the biological macromolecule by IR-MALDI. For example, an additive can improve the miscibility of the biological macromolecule in the liquid matrix. For example, a composition can contain a nucleic acid as the biological macromolecule to be analyzed by IR-MALDI and glycerol as the liquid matrix. The liquid matrix can be treated with a cation exchange material prior to mixing

with the nucleic acid, if desired, to reduce alkali salt formation with the phosphate backbone.

A composition for IR-MALDI can deposited on a substrate, for example, a solid support such as a silicon wafer, a bead, other support know to those of skill in the art, thereby providing a solid support having deposited thereon a composition for IR-MALDI.

In particular, the solid support can be a silicon wafer and a plurality of compositions for IR-MALDI can be deposited on the wafer in an addressable array. If desired, a composition for IR-MALDI can contain two or more different biological macromolecules to be analyzed, provided the biological macromolecules are differentially identifiable due, for example, to mass modification.

Liquid matrices

As defined above, a liquid matrix refers to a material that is compatible with the macromolecule of interest, absorbs IR, and can form a glass (rather than a crystalline structure). A liquid matrix has a sufficient absorption at the wavelength of the laser to be used in performing desorption and ionization and is a liquid (not a solid or a gas) at room temperature (one atmosphere pressures).

In addition, for purposes herein in performing IR-MALDI, contemplated matrices in embodiments for methods of diagnosis and detection of proteins and nucleic acids also can include materials that form crystalline structures. Such materials include, but are not limited to, water, ice and succinic acid and piccolinic acid and other acids. These types of materials include those that do form ordered structures when cooled, dried and/or are under pressure. These types of matrices are contemplated for use in detection methods of proteins using IR MALDI. When succinic acid is dipensed on a selected substrate (or support) for IR MALDI, preferably, nucleic acid should be added prior to dispensing. For other matrices that are dried on the the substrate, nucleic acids can be added to the dried matrix material.

For absorption purposes, the liquid matrix can contain at least one chromophore or functional group that strongly absorbs infrared radiation. Examples of appropriate functional groups include nitro, sulfonyl, sulfonic acid, sulfonamide, nitrile or cyanide, carbonyl, aldehyde, carboxylic acid, amide, ester, anhydride, ketone, amine, hydroxyl, aromatic rings, dienes and other conjugated systems.

Preferred liquid matrices, include but are not limited to, substituted or unsubstituted (1) alcohols, preferably non-volatile liquids (or liquids of low volatility), including glycols, such as glycerol, 1,2-propanediol or 1,3-propanediol, 1,2-butanediol, 1,3-butanediol, 1,4-butanediol and triethanolamine, sucrose, mannose and other polyols; (2) carboxylic acids including formic acid, lactic acid, acetic acid, propionic acid, butanoic acid, pentanoic acid and hexanoic acid, and esters thereof; (3) primary or secondary amides, including acetamide, propanamide, butanamide, pentanamide and hexanamide, whether branched or unbranched; (4) primary or secondary amines, including propylamine, butylamine, pentylamine, hexylamine, heptylamine, diethylamine and dipropylamine; (5) nitriles, hydrazine and hydrazide.

Particularly preferred compounds contain eight or fewer carbon atoms. For example, particularly preferred carboxylic acids and amides contain six or fewer carbon atoms, preferred amines contain about three to about seven carbons and preferred nitriles contain eight or fewer carbons. Compounds that are unsaturated to any degree can contain a larger number of carbons, since unsaturation confers liquid properties on a compound. Although the particular compound used as a liquid matrix must contain a functional group, the matrix preferably is not so reactive that it fragments or otherwise damages the nucleic acid to be analyzed.

An appropriate liquid matrix should be miscible with a nucleic acid compatible solvent. Preferably, the liquid matrix also should have an appropriate viscosity, for example, typically less than or equal to about 1.5 s/m^2, preferably in the range of about 1 s/m^2 to about 2 s/m^2, which is the viscosity of glycerol at room temperature, to facilitate dispensing of microliter or

nanoliter volumes of matrix alone or mixed with a nucleic acid compatible solvent.

For use herein, a liquid matrix also should have an appropriate survival time in the vacuum of the analyzer, typically having a pressure in the range of about 10^{-10} mbars, to allow the analysis to be completed. Liquids having an appropriate survival time are "vacuum stable," a property that is strictly a function of the vapor pressure of the matrix, which, in turn, is strongly dependent on the sample temperature. Preferred matrices have a low vapor pressure at room temperature such that less than about fifty percent of the sample in a mass analyzer having a back pressure less than or equal to 10^{-5} mbars evaporates in the time needed for the analysis of all samples introduced, for example, about 10 minutes to about 2 hours. For a single sample, for example, the analysis may be performed in minutes, whereas, for multiple samples, the analysis may require hours for completion.

Glycerol, for example, can be used as a matrix at room temperature and in a vacuum for about 10 to 15 minutes. If glycerol is to be used for analyzing multiple samples in a single vacuum, the vacuum may need to be cooled to maintain the sample at a temperature in the range of about -50°C to about -100°C (about 173°K to about 223°K) for the time required to complete the analysis. Colder temperatures can also be used, including as low as about -200° C. Triethanolamine, in contrast, has a much lower vapor pressure than glycerol and can survive in a vacuum for at least about one hour, even at room temperature.

Mixtures of different liquid matrices and additives to such matrices may be desirable to confer one or more of the properties described above. For example, an appropriate liquid matrix can contain a small amount of a composition containing an IR absorbing chromophore and a greater amount of an IR invisible (nonabsorbing) material, in which, for example, the nucleic acid is soluble. It also may be useful to use a matrix that is "doped" with a small amount of a compound or compounds having a high extinction coefficient (E) at the laser wavelength used for desorption and ionization, for example, dinitrobenzenes or polyenes. An additive that acidifies the liquid matrix also

may be added to dissociate double stranded nucleic acids or to denature secondary structure of nucleic acids such as tRNA or other RNA. Additional additives may be helpful for minimizing salt formation between the matrix and the phosphate backbone of the nucleic acid. For example, the additive can contain an ammonium salt or ammonium loaded ion exchange bead, which removes alkali ions from the matrix. Alternatively, the liquid matrix can be distilled prior to mixture with the nucleic acid composition, to minimize salt formation between the matrix and the phosphate backbone of the nucleic acid.

The liquid matrix also can be mixed with an appropriate volume of water or other liquid to control sample viscosity and rate of evaporation. Since all of the water is evaporated during mass analysis, an easily manipulated volume, for example, 1 μl, can be useful for sample preparation and transfer, but still result in a very small volume of liquid matrix. As a result, only small volumes of nucleic acid sample are required to yield about 10^{-16} to about 10^{-12} moles (about 100 attomol to about 1 pmol) of nucleic acid in the final liquid matrix droplet.

As disclosed herein, when glycerol is used as a matrix, the final analyte-to-glycerol molar ratio (concentration) should be in the range of about 10^{-4} to 10^{-9}, depending on the mass of the nucleic acid, which can range up to about 10^4 Daltons to about 10^6 Daltons or greater, and the total amount of nucleic acid available. For example, for the sensitivity test disclosed herein, the relatively high concentration of nucleic acid used was measured by standard UV spectrophotometry. Practically speaking, the appropriate amount of nucleic acid generated, for example, from a PCR or transcription reaction generally is known. The large range specified indicates that the actual amount of nucleic acid analyzed is not very critical. Typically, a greater amount of nucleic acid results in a better spectrum. There may be instances where the nucleic acid sample requires dilution.

Other liquid matrices include, but are not limited to triethanolamine, lactic acid, 3-nitrobenzylalcohol, diethanolamine, DMSO, nitropheynloctylether (3-NPOE), 2,2'dithiodiethanol, tetraethyleneglycol, dithiotrietol/erythritol (DTT/DTE), 2,3-dihydroxy-propyl-benzyl ether, α-tocopherol, and thioglycerol.

IMMOBILIZATION OF A BIOLOGICAL MACROMOLECULE TO A SOLID SUPPORT OR SUBSTRATE

For IR-MALDI mass spectrometric analyses, a target biological macromolecule or other biological macromolecule of interest can be immobilized to a substrate, particularly a solid support, in order to facilitate manipulation of the biological macromolecule. Solid supports are well known in the art and include any material used as a solid support for linking nucleic acids, proteins, carbohydrates, or the like (see, for example, International Publ. WO 98/20019).

The substrate can be selected to be impervious to the conditions of IR-MALDI mass spectrometric analyses, and can be functionalized for the immobilization of biological macromolecules or can be further associated with a second solid support, if desired. Where a substrate, for example, a bead is to be conjugated to a second solid support, biological macromolecules can be immobilized on the functionalized bead before, during or after it is conjugated to the second support.

A biological macromolecule can be conjugated directly to a solid support or can be immobilized indirectly through a functional group present either on the support, or a linker attached to the support, or the biological macromolecule or both. For example, a polypeptide can be immobilized to a solid support through a hydrophobic, hydrophilic or ionic interaction between the support and the polypeptide. Although such a method can be useful for certain manipulations such as for conditioning of the biological macromolecule prior to IR-MALDI mass spectrometry, such a direct interaction is limited in that the orientation of the biological macromolecule is not known and can be random based on the position of the interacting subunits, for example, hydrophobic amino acids in a polypeptide. Thus, a polypeptide or other biological macromolecule generally is immobilized in a defined orientation by conjugation through a functional group on either the solid support or the biological macromolecule or both.

A biological macromolecule can be modified by adding an appropriate functional group to a terminus of the biological macromolecule, for example, to the 5′ or 3′ end of a nucleic acid, or to the carboxyl terminus or amino terminus of a polypeptide, or to a reactive group in the biological macromolecule, for

example, to a reactive group of a nucleotide or to the phosphodiester backbone of a nucleic acid, or to a reactive side chain of an amino acid or to the peptide backbone of a polypeptide. A naturally occurring nucleotide in a nucleic acid or a naturally occurring amino acid in a polypeptide also can contain a functional group suitable for conjugating the polypeptide to the solid support. For example, a cysteine residue present in the polypeptide can be used to immobilize the polypeptide to a substrate containing a sulfhydryl group, for example, a solid support having cysteine residues attached thereto, through a disulfide linkage. Other bonds that can be formed between two amino acids, for example, include monosulfide bonds between two lanthionine residues, which are non-naturally occurring amino acids that can be incorporated into a polypeptide; a lactam bond formed by a transamidation reaction between the side chains of an acidic amino acid and a basic amino acid, such as between the γ-carboxyl group of Glu (or β-carboxyl group of Asp) and the ϵ-amino group of Lys; or a lactone bond produced, for example, by a crosslink between the hydroxy group of Ser and the γ-carboxyl group of Glu (or β-carboxyl group of Asp). Thus, a solid support can be modified to contain a desired amino acid residue, for example, a Glu residue, and a polypeptide having a Ser residue, particularly a Ser residue at the carboxyl terminus or amino terminus, can be conjugated to the solid support through the formation of a lactone bond. It should be recognized, however, that the support need not be modified to contain the particular amino acid, for example, Glu, where it is desired to form a lactone-like bond with a Ser in the polypeptide, but can be modified, instead, to contain an accessible carboxyl group, thus providing a function corresponding to the γ-carboxyl group of Glu.

A biological macromolecule can be modified to facilitate immobilization to a solid support, for example, by incorporating a chemical or physical moiety at an appropriate position in the biological macromolecule, generally at a terminus of the biological macromolecule. The artisan will recognize, however, that such a modification, for example, the incorporation of a biotin moiety, can affect the ability of a particular reagent to interact specifically with the biological

macromolecule and, accordingly, will consider this factor, if relevant, in selecting how best to modify a biological macromolecule of interest.

In one aspect of the processes provided herein, a polypeptide of interest can be covalently conjugated to a solid support and the immobilized polypeptide can be used to capture a target polypeptide, which binds to the immobilized polypeptide. The target polypeptide then can be released from immobilized polypeptide by ionization or volatization for IR-MALDI mass spectrometry, whereas the covalently conjugated polypeptide remains bound to the support.

Accordingly, a process as disclosed herein can utilize IR-MALDI to determine the identity of polypeptides that interact specifically with a polypeptide of interest. For example, the identity of target polypeptides obtained from one or more biological samples that interact specifically with a immobilized polypeptide of interest can be determined, or the identity of binding proteins such as antibodies that bind to the immobilized polypeptide antigen of interest, or receptors that bind to an immobilized polypeptide ligand of interest, or the like can be determined. Such a process can be useful, for example, for screening a combinatorial library of modified target polypeptides such as modified antibodies, antigens, receptors, hormones, or other polypeptides to determine the identity of those target polypeptides that interact specifically with the immobilized polypeptide.

A solid support can be selected based on advantages that it can provide. For example, a solid support can provide a relatively large surface area, thereby allowing immobilization of a relatively large number of biological macromolecules. A solid support such as a bead can have any three dimensional structure, including a surface to which a biological macromolecule, functional group, or other molecule can be attached.

A substrate also can be modified to facilitate immobilization of a biological macromolecule. A thiol-reactive functionality is particularly useful for immobilizing a polypeptide to a solid support (International Publ. WO 98/20166). A thiol-reactive functionality can rapidly react with a nucleophilic thiol moiety to produce a covalent bond, for example, a disulfide bond or a thioether bond. A variety of thiol-reactive functionalities are known in

the art, including, for example, haloacetyls such as iodoacetyl; diazoketones; epoxy ketones; α- and ß-unsaturated carbonyls such as α-enones and ß-enones; and other reactive Michael acceptors such as maleimide; acid halides; benzyl halides; and the like. A free thiol group of a disulfide, for example, can react with a second free thiol group by disulfide bond formation, including by disulfide exchange. Reaction of a thiol group or other functional group can be prevented temporarily by blocking with an appropriate protecting group (see Greene and Wuts, *Protective Groups in Organic Synthesis* 2nd ed. (John Wiley & Sons 1991)).

A thiol-reactive functionality such as 3-mercaptopropyltriethoxysilane can be used to functionalize a silicon surface with thiol groups. The amino functionalized silicon surface then can be reacted with a heterobifunctional reagent such as N-succinimidyl (4-iodacetyl) aminobenzoate (SIAB; Pierce; Rockford IL). If desired, the thiol groups can be blocked with a photocleavable protecting group, which then can be selectively cleaved, for example, by photolithography, to provide portions of a surface activated for immobilization of a polypeptide of interest. Photocleavable protecting groups are known in the art (see, for example, International Publ. WO 92/10092; McCray et al., Ann. Rev. Biophys. Biophys. Chem. 18:239-270 (1989)) and can be selectively deblocked by irradiation of selected areas of the surface using, for example, a photolithography mask.

Solid Supports (substrates)

The solid support is any known to those of skill in the art as matrix for performing synthetic reactions and assays. It can be fabricated from silicon, glass, silicon-coated materials, metal, a composite, a polymeric material such as a plastic, a polymer-grafted material, suich as a metal-grafted polymer, or other material as disclosed herein. This material can be further functionalized, as necessary, for example, chemically, to enhance or permit linkage of molecules or other particles, such as cells or cell membranes or viral envelopes or other such biological materials, of interest. The surface of a support can be modified, such as by radiation grafting of a suitable polymer on the surface and derivatization thereof to render it suitable for binding capturing a molecule or

particle, such as a cell. The support may also include beads linked thereto (see, copending allowed U.S. application Serial No. 08/746,036, copending U.S. application Serial No. 08/933,792, and International application No. PCT/US97/20194, which claims priority to the U.S. applications). It may also include dendrite trees of captured material, or combinations of such additional components. A solid support can have one or more target sites, each of which can contain or retain a volume of a liquid.

By way of example, a solid support can be a flat surface such as a glass fiber filter, a glass surface, a silicon or silicon dioxide surface, a composite surface, or a metal surface, including a steel, gold, silver, aluminum or copper surface, a plastic material, including polyethylene, polypropylene, polyamide or polyvinylidenedifluoride, which further can be in the form of multiwell plate or a membrane; can be in the form of a bead (or other geometry) or particle, such as a silica gel, a controlled pore glass, a magnetic or cellulose bead, which can be in a pit of a flat surface such as a wafer, for example, a silicon wafer; or can be a pin, including an array of pins suitable for combinatorial synthesis or analysis (see, *e.g.*, International PCT application No. WO98/20019), comb, microchip. The skilled artisan will recognize that various factors, including the size and shape of the support and the chemical and physical stability of the support to the conditions to which it will be exposed, will be considered in selecting a particular solid support for use in a disclosed system or method.

Also contemplated is the use of the end of a fiber optic cable or plate as a substrate or support (see, *e.g.*, U.S. Patent No. 5,826,214, which describes embodiments in which the electromagnetic radiation is delivered via a fiber optic cable, which can abut against a thin transparent plate on which the specimen or resides).

A solid support contains one or more target sites, which can contain a volume of a liquid. A target site can be, for example, a well, pit, channel, or other depression, with or without rims, on the surface of a solid support; can be a pin, bead or other material, which can be positioned on a surface of a solid support; or can be a physical barrier such as a cylinder, cone or other such barrier positioned on a surface of a solid support.

A target site also can be, for example, a reservoir or reaction chamber, which is attached to a solid support (see, for example, Walters *et al.*, Anal. Chem. 70:5172-5176 (1998)). In addition, a target site can be etched, for example, on a surface of a silicon wafer using a photolithographic method (see, for example, Woolley *et al.* (Anal. Chem. 68:4081-4086 (1996)). Photolithography allows the construction of very small target sites, including wells or towers, and, for example, has been used in combination with wet chemical-etching to construct "picoliter vials" on microchips (Clark *et al.* CHEMTECH 28:20-25 (1998)).

A support also can be a glass or silicon surface containing wells having a very thin base that is transparent to electromagnetic radiation of a desired wavelength, such as laser light, thereby permitting measurement of parameters, such as volume, or an excitation wavelength for fluorescence measurement.

A target site also can be defined by physico-chemical parameters such as hydrophilicity, hydrophobicity, the presence of acidic or basic groups, groups capable of forming a salt bridge, or any surface chemistry that allows a liquid to grow primarily in the z direction. For example, where the liquid to be placed on a target site is water or an aqueous composition, the target site can be defined by a hydrophilic area surrounded by a hydrophobic area on the surface of a solid support, or by a series of rows, alternately having less hydrophobic rows and more hydrophobic rows, whereby the aqueous mixture is constrained to the less hydrophobic rows. With respect to such a target site, the aqueous composition is dispensed, for example, onto the hydrophilic area, and is constrained from spreading from the target site due to the adjacent and surrounding hydrophobic area. Conversely, where the liquid is a nonpolar liquid, it is dispensed onto a hydrophobic region and is constrained in that region due to an adjacent hydrophilic region or a region or that is less hydrophobic that the region to which the liquid is applied.

A solid support can have a single target site, or can contain a number of target sites, for example, 2 sites, 10 sites, 16 sites, 100 sites, 144 sites, 384 sites, 1000 sites, or more, all or some of which can be the same or can be different. Where a solid support contains more than one target site and,

therefore, can contain, for example, more than one reaction mixture, the characteristics that define each target site serve not only to constrain a reaction mixture, but also to prevent intermingling of different reaction mixtures or other liquids on the support. In addition, where a solid support contains more than one target site, the target sites can be arranged in any pattern, for example, in a line, a spiral, concentric circles, rows, or an array of rows and columns. Furthermore, the location of each target site of a number of target sites on a support can be defined. The availability of such addressable target sites on a solid support allows multiple reactions to be performed in parallel and is convenient, for example, for performing multiplex reactions, for including control reactions with test reactions such that all are performed under identical conditions, for performing a similar reaction under different conditions, or for performing different reactions.

Thus, any substrate on which the nucleic acid/liquid matrix can be deposited and retained for desorption and ionization of the nucleic acid can be used in a process provided herein. Preferred substrates include, but are not limited to beads, for example, silica gel, controlled pore glass, magnetic, cross-linked dextrans, such as those sold under the tradename Sephadex (Pharmacia) and agarose gel, such as gels sold under the tradename Sepharose (Pharmacia), which is a hydrogen bonded polysaccharide-type agarose gel (epichlorhydrins), or cellulose; capillaries; flat supports, for example, filters, plates or membranes made of glass, metal surfaces such as steel, gold, silver, aluminum, copper or silicon, or plastic such as polyethylene, polypropylene, polyamide or polyvinylidene fluoride; pins, for example, arrays of pins suitable for combinatorial synthesis or analysis of beads in pits of flat surfaces such as wafers, with or without filter plates.

Preferably the selected substrate and format are amenable to miniaturization, such as the chips that retain the deposited material by virue of hydrophobic or hydrophilic interaction, described above, in which the target site can be defined by a hydrophilic area surrounded by a hydrophobic area on the surface of a solid support (or the converse).

Preferably, nucleic acid samples are prepared and deposited as a thin layer, for example, a monolayer to about a 100 μm layer, preferably between about 0.1 μm and about 100 μm, more preferably 1 μm to 10 μm, onto a substrate manually or using an automated device, so that multiple samples can be prepared and analyzed on a single sample support plate with only one transfer into the vacuum of the analyzer and requiring only a relatively short period of time for analysis. Appropriate automated sample handling systems for use in the instant process are described, for example, in U.S. Patent Nos. 5,705,813; 5,716,825; and 5,498,545 and co-pending U.S. application Serial No. 09/285,481, as well as allowed U.S. application Serial No. 08/787,639, and published International PCT application WO 98/20166.

Immobilization and activation

Numerous methods have been developed for the immobilization of proteins, nucleic acids and other biomolecules onto solid or liquid supports [see, e.g., Mosbach (1976) Methods in Enzymology 44; Weetall (1975) Immobilized Enzymes, Antigens, Antibodies, and Peptides; and Kennedy et al. (1983) Solid Phase Biochemistry, Analytical and Synthetic Aspects, Scouten, ed., pp. 253-391; see, generally, Affinity Techniques. Enzyme Purification: Part B. Methods in Enzymology, Vol. 34, ed. W. B. Jakoby, M. Wilchek, Acad. Press, N.Y. (1974); Immobilized Biochemicals and Affinity Chromatography, Advances in Experimental Medicine and Biology, vol. 42, ed. R. Dunlap, Plenum Press, N.Y. (1974)].

Among the most commonly used methods are absorption and adsorption or covalent binding to the support, either directly or via a linker, such as the numerous disulfide linkages, thioether bonds, hindered disulfide bonds, and covalent bonds between free reactive groups, such as amine and thiol groups, known to those of skill in art [see, e.g., the PIERCE CATALOG, ImmunoTechnology Catalog & Handbook, 1992-1993, which describes the preparation of and use of such reagents and provides a commercial source for such reagents; and Wong (1993) Chemistry of Protein Conjugation and Cross Linking, CRC Press; see, also DeWitt et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90:6909; Zuckermann et al. (1992) J. Am. Chem. Soc. 114:10646; Kurth et al.

(1994) J. Am. Chem. Soc. 116:2661; Ellman et al. (1994) Proc. Natl. Acad. Sci. U.S.A. 91:4708; Sucholeiki (1994) Tetrahedron Lttrs. 35:7307; and Su-Sun Wang (1976) J. Org. Chem. 41:3258; Padwa et al. (1971) J. Org. Chem. 41:3550 and Vedejs et al. (1984) J. Org. Chem. 49:575, which describe photosensitive linkers]

To effect immobilization, a composition of the protein or other biomolecule is contacted with the support material such as any described herein, alumina, carbon, an ion-exchange resin, cellulose, glass or a ceramic. Fluorocarbon polymers have been used as supports to which biomolecules have been attached by adsorption [see, U.S. Pat. No. 3,843,443; Published International PCT Application WO/86 03840].

A large variety of methods are known for attaching biological molecules, including proteins and nucleic acids, molecules to solid supports [see. e.g., U.S. Patent No. 5451683]. Such linkages may be effected through covalent bonds, ionic bonds and other interactions. The linkages may be reversible or labile to certain conditions, such as particular EM frequencies.

For example, U.S. Pat. No. 4,681,870 describes a method for introducing free amino or carboxyl groups onto a silica support. These groups may subsequently be covalently linked to other groups, such as a protein or other anti-ligand, in the presence of a carbodiimide. Alternatively, a silica support may be activated by treatment with a cyanogen halide under alkaline conditions. The anti-ligand is covalently attached to the surface upon addition to the activated surface. Another method involves modification of a polymer surface through the successive application of multiple layers of biotin, avidin and extenders [see, e.g., U.S. Patent No. 4,282,287]; other methods involve photoactivation in which a polypeptide chain is attached to a solid substrate by incorporating a light-sensitive unnatural amino acid group into the polypeptide chain and exposing the product to low-energy ultraviolet light [see, e.g., U.S. Patent No. 4,762,881].

Oligonucleotides have also been attached using a photochemically active reagents, such as a psoralen compound, and a coupling agent, which attaches the photoreagent to the substrate [see, e.g., U.S. Patent No. 4,542,102 and

U.S. Patent No. 4,562,157]. Photoactivation of the photoreagent binds a nucleic acid molecule to the substrate to give a surface-bound probe.

Covalent binding of the protein or other biomolecule or organic molecule or biological particle to chemically activated solid support supports such as glass, synthetic polymers, and cross-linked polysaccharides is a more frequently used immobilization technique. The molecule or biological particle may be directly linked to the support or linked via linker, such as a metal [see, e.g., U.S. Patent No. 4,179,402; and Smith et al. (1992) Methods: A Companion to Methods in Enz. 4:73-78]. An example of this method is the cyanogen bromide activation of polysaccharide supports, such as agarose. The use of perfluorocarbon polymer-based supports for enzyme immobilization and affinity chromatography is described in U.S. Pat. No. 4,885,250]. In this method the biomolecule is first modified by reaction with a perfluoroalkylating agent such as perfluorooctylpropylisocyanate described in U.S. Pat. No. 4,954,444. Then, the modified protein is adsorbed onto the fluorocarbon support to effect immobilization.

The activation and use of supports are well known and may be effected by any such known methods [see, e.g., Hermanson et al. (1992) Immobilized Affinity Ligand Techniques, Academic Press, Inc., San Diego]. For example, the coupling of the amino acids may be accomplished by techniques familiar to those in the art and provided, for example, in Stewart and Young, 1984, Solid Phase Synthesis, Second Edition, Pierce Chemical Co., Rockford.

Molecules may also be attached to supports through kinetically inert metal ion linkages, such as Co(III), using, for example, native metal binding sites on the molecules, such as IgG binding sequences, or genetically modified proteins that bind metal ions [see, e.g., Smith et al. (1992) Methods: A Companion to Methods in Enzymology 4, 73 (1992); Ill et al. (1993) Biophys J. 64:919; Loetscher et al. (1992) J. Chromatography 595:113-199; U.S. Patent No. 5,443,816; Hale (1995) Analytical Biochem. 231:46-49].

Other suitable methods for linking molecules and biological particles to solid supports are well known to those of skill in this art [see, e.g., U.S. Patent No. 5,416,193]. These linkers include linkers that are suitable for chemically linking molecules, such as proteins and nucleic acid, to supports include, but are not limited to, disulfide bonds, thioether bonds, hindered disulfide bonds, and covalent bonds between free reactive groups, such as amine and thiol groups. These bonds can be produced using heterobifunctional reagents to produce reactive thiol groups on one or both of the moieties and then reacting the thiol groups on one moiety with reactive thiol groups or amine groups to which reactive maleimido groups or thiol groups can be attached on the other. Other linkers include, acid cleavable linkers, such as bismaleimideothoxy propane, acid labile-transferrin conjugates and adipic acid diihydrazide, that would be cleaved in more acidic intracellular compartments; cross linkers that are cleaved upon exposure to UV or visible light and linkers, such as the various domains, such as C_H1, C_H2, and C_H3, from the constant region of human IgG_1 (see, Batra et al. (1993) Molecular Immunol. 30:379-386). Presently preferred linkages are direct linkages effected by adorbing the molecule or biological particle to the surface of the support. Other preferred linkages are photocleavable linkages that can be activated by exposure to light [see, e.g., Goldmacher et al. (1992) Bioconj. Chem. 3:104-107, which linkers are herein incorporated by reference]. The photocleavable linker is selected such that the cleaving wavelength that does not damage linked moieties. Photocleavable linkers are linkers that are cleaved upon exposure to light [see, e.g., Hazum et al. (1981) in Pept., Proc. Eur. Pept. Symp., 16th, Brunfeldt, K (Ed), pp. 105-110, which describes the use of a nitrobenzyl group as a photocleavable protective group for cysteine; Yen et al. (1989) Makromol. Chem 190:69-82, which describes water soluble photocleavable copolymers, including hydroxypropylmethacrylamide copolymer, glycine copolymer, fluorescein copolymer and methylrhodamine copolymer; Goldmacher et al. (1992) Bioconj. Chem. 3:104-107, which describes a cross-linker and reagent that undergoes photolytic degradation upon exposure to near UV light (350 nm); and Senter et al. (1985) Photochem. Photobiol 42:231-237, which describes nitrobenzyloxycarbonyl chloride cross linking reagents that

produce photocleavable linkages]. The selected linker will depend upon the particular application and, if needed, may be empirically selected.

Linkers

A biological macromolecule can be immobilized directly to a substrate or can be immobilized through a linking moiety or moieties. Immobilization can be effected by any desired linkage including covalent linkages, ionic linkages, physical linkages, and any other linkages known. The linkage can be reversible and/or cleavable. Any linker known to those of skill in the art to be suitable for immobilizing a nucleic acid, polypeptide, carbohydrate or other biological macromolecule to a substrate, either directly or through a spacer, can be used (see International Publ. WO 98/20019). Among preferred linkers are those that are cleave or otherwise release upon exposure to IR.

A biological macromolecule can be immobilized directly to a support through a linker or can be immobilized through a variable spacer. In addition, the conjugation can be directly cleavable, for example, through a photocleavable linkage such as a streptavidin or avidin to biotin interaction, which can be cleaved by a laser, or indirectly through a photocleavable linker (U.S. Patent No. 5,643,722) or an acid labile linker, heat sensitive linker, enzymatically cleavable linker or other such linker. Accordingly, a linker can provide a reversible linkage such that it is cleaved under defined conditions such as during the IR-MALDI mass spectrometry procedure. Such a linker can be, for example, a photocleavable bond such as a charge transfer complex or a labile bond formed between relatively stable organic radicals.

A linker (L) on a biological macromolecule can form a linkage, which generally is a temporary linkage, with a second functional group (L') on the solid support. Furthermore, where the biological macromolecule has a net negative charge, or is conditioned to have such a charge, the linkage can be formed with L' being, for example, a quaternary ammonium group. In this case, the surface of the solid support carries a negative charge that repels the negatively charged biological macromolecule, thereby facilitating desorption of the biological macromolecule for IR-MALDI mass spectrometric analysis. Desorption can

occur due to the heat created by the IR radiation or, where L' is a chromophore, by specific absorption of IR radiation by the chromophore.

A linkage (L-L') can be, for example, a disulfide bond, which is chemically cleavable by mercaptoethanol or dithioerythrol; a biotin/streptavidin linkage, which can be photocleavable; a heterobifunctional derivative of a trityl ether group, which can be cleaved by exposure to acidic conditions (see Köster et al., Tetrahedron Lett. 31:7095 (1990)); a levulinyl-mediated linkage, which can be cleaved under almost neutral conditions with a hydrazinium/acetate buffer; an arginine-arginine or a lysine-lysine bond, either of which can be cleaved by an endopeptidase such as trypsin; a pyrophosphate bond, which can be cleaved by a pyrophosphatase; or a ribonucleotide bond, which can be cleaved using a ribonuclease or by exposure to alkali condition.

The functionalities, L and L', can also form a charge transfer complex, thereby forming a temporary L-L' linkage. The IR laser energy can be tuned to the corresponding energy of the charge-transfer wavelength and specific desorption from the solid support can be initiated. It will be recognized that several combinations of L and L' can serve this purpose and that the donor functionality can be on the solid support or can be coupled to the biological macromolecule to be detected or vice versa, provided a liquid matrix, which absorbs IR radiation, also is present.

Selectively cleavable linkers that are particularly useful in a process as disclosed herein include photocleavable linkers, acid cleavable linkers, acid-labile linkers, and heat sensitive linkers. Acid cleavable linkers include, for example, bis-maleimideothoxy propane, adipic acid dihydrazide linkers (Fattom et al., Infect. Immun. 60:584-589 (1992)), and acid labile transferrin conjugates that contain a sufficient portion of transferrin to permit entry into the intracellular transferrin cycling pathway (Welhöner et al., J. Biol. Chem. 266:4309-4314

(1991)). Photocleavable linkers also include the linkers described in WO 98/20019.

Linkers suitable for chemically linking polypeptides, for example, to supports, include disulfide bonds, thioether bonds, hindered disulfide bonds, and covalent bonds between free reactive groups such as amine and thiol groups.

Agents useful for creating linkages include, for example, dimaleimide, dithio-bis-nitrobenzoic acid (DTNB), N-succinimidyl-S-acetyl-thioacetate (SATA), N-succinimidyl-3-(2-pyridyldithiol propionate (SPDP), succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) 6-hydrazino nicotimide (HYNIC). Appropriate linkers, which can be crosslinking agents, for use for conjugating a polypeptide to a solid support include a variety of agents that can react with a functional group present on a surface of the support, or with the polypeptide, or both. Useful crosslinking agents include agents containing homobifunctional or heterobifunctional groups. Useful bifunctional crosslinking agents include, but are not limited to, N-succinimidyl(4-iodoacetyl) aminobenzoate (SIAB), dimaleimide, dithio-bis-nitrobenzoic acid (DTNB), N-succinimidyl-S-acetyl-thioacetate (SATA), N-succinimidyl-3-(2-pyridyldithio) propionate (SPDP), succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) and 6-hydrazino-nicotimide (HYNIC).

A crosslinking agent also can be used to form a selectively cleavable bond between a biological macromolecule and a solid support. For example, a photolabile crosslinker such as 3-amino-(2-nitrophenyl)propionic acid (Brown et al., Molec. Divers. 4-12 (1995); Rothschild et al., Nucl. Acids Res. 24:351-66 (1996); U.S. Patent No. 5,643,722) can be employed as a means for cleaving a polypeptide from a solid support. Other crosslinking reagents are well known in the art (see, for example, Wong, *Chemistry of Protein Conjugation and Cross-Linking* (CRC Press 1991); Hermanson, *Bioconjugate Techniques* (Academic Press 1996)).

D Hydroxyester linkers, including, for example, hydroxyacetate (glycolate), α-, β-, γ-,..., ω-hydroxyalkanoates, ω-hydroxy(polyethylene glycol)COOH, hydroxybenzoates, hydroxyarylalkanoates and hydroxyalkylbenzoates, can be useful for immobilizing a biological macromolecule. Photocleavable linkers also

are useful for immobilizing a biological macromolecule; methods of preparing such linkers are provided in International Publ. WO 98/20019. In addition, a bifunctional trityl linker can be attached to a solid support, for example, to the 4-nitrophenyl active ester on a resin such as a Wang resin, through an amino group or a carboxyl group on the resin via an amino resin. Using a bifunctional trityl approach, the solid support can require treatment with a volatile acid such as formic acid or trifluoracetic acid to ensure that the biological macromolecule can be removed. In such a case, the biological macromolecule can be deposited as a beadless patch at the bottom of a well of a solid support or on the flat surface of a solid support. After addition of a matrix composition, the biological macromolecule can be desorbed during IR-MALDI mass spectrometry.

Hydrophobic trityl linkers also can be exploited as acid-labile linkers by using a volatile acid or an appropriate matrix composition, which is acidic or contains an additive that renders the liquid matrix acidic, to cleave an amino linked trityl group from the biological macromolecule. Acid lability also can be changed. For example, trityl, monomethoxytrityl, dimethoxytrityl or trimethoxytrityl can be changed to the appropriate p-substituted, or more acid-labile tritylamine derivatives.

Other linkers, include, for example, Rink amide linkers (Rink, Tetrahedron Letters 28:3787 (1976)), tritylchloride linkers (Leznoff, Ace. Chem. Res. 11:327 (1978)), Merrifield linkers (Bodansky et al., Peptide Synthesis 2d ed., Academic Press; New York, 1986); trityl linkers (U.S. Patent Nos. 5,410,068 and 5,612,474); and amino trityl linkers (U.S. Patent No. 5,198,531).

Other linkers include acid cleavable linkers such as bis-maleimideothoxy propane, acid labile transferrin conjugates and adipic acid dihydrazide linkers that can be cleaved in more acidic intracellular compartments; photocleavable cross linkers that are cleaved by IR, visible or UV light, RNA linkers that are cleavable by ribozymes or other RNA enzymes, and linkers such as the various domains, including C_H1, C_H2, and C_H3, from the constant region of human IgG_1 (see, Batra et al., Mol. Immunol. 30:379-386 (1993)). Combinations of any linkers also can be useful, for example, a linker that can be cleavable under IR-MALDI mass spectrometric conditions such as a silyl linkage or photocleavable

linkage can be combined with a linker such as an avidin biotin linkage, which is not cleaved under IR-MALDI mass spectrometry conditions but can be cleaved under other conditions.

A biological macromolecule of interest can be immobilized to a solid support such as a bead. In addition, a first solid support such as a bead also can be conjugated to a second solid support, which can be a second bead or other substrate, by any suitable means. In particular, any of the conjugation methods and means disclosed herein with reference to conjugation of a biological macromolecule to a solid support also can be applied for conjugation of a first support to a second support, where the first and second solid supports can be the same or different. Furthermore, use of bifunctional linkers allows for orthogonal cleavage of a biological macromolecule from a support, or of a first support from a second.

It should be recognized that any of the binding members disclosed herein or otherwise known in the art can be reversed with respect to the examples provided herein. Thus, biotin, for example, can be incorporated into either a biological macromolecule or a solid support and, conversely, avidin or other biotin binding moiety would be incorporated into the support or the polypeptide, respectively. Other specific binding pairs contemplated for use herein are exemplified by hormones and their receptors, enzymes and their substrates, a nucleotide sequence and its complementary sequence, an antibody and the antigen to which it interacts specifically, and other such pairs known to those skilled in the art.

A target biological macromolecule, particularly each target biological macromolecule in a plurality of target biological macromolecules, can be immobilized to a solid support prior to mass modifying, conditioning, or otherwise manipulating the biological macromolecule. In particular, the solid support can be a flat surface, or a surface with a structure such as wells, such that each of the target biological macromolecules in the plurality can be positioned in an array, each at a particular address. In general, a target biological macromolecule is immobilized to the solid support through a cleavable linker such as an acid labile linker, a chemically cleavable linker or a

photocleavable linker. Following a reaction of the target biological macromolecule in a disclosed process, undesirable reaction products can be washed from the reaction and the remaining immobilized target biological macromolecule can be released, for example, by chemical cleavage or photocleavage, as appropriate, and can be analyzed by IR-MALDI mass spectrometry. It should be recognized, however, that manipulation of a biological macromolecule, for example, by mass modification prior to performing a chemical or enzymatic degradation or other reaction can influence the rate or extent of the reaction. Accordingly, the skilled artisan will know that the influence of conditioning, mass modification, or the like on the extent of a reaction should be characterized prior to initiating a process.

In some cases, it can be useful to immobilize a particular target biological macromolecule to a support through both termini of the biological macromolecule, for example, the amino terminus and the carboxyl terminus of a polypeptide using, for example, a chemically cleavable linker at one terminus and a photocleavable linker at the other end. In this way, the target biological macromolecule, which can be immobilized, for example, in an array in wells, can be contacted, for example, with one or more agents that cleave at least one bond linking the monomer subunits in the biological macromolecule, the internal biological macromolecule fragments then can be washed from the wells, along with the agent and any reagents in the well, leaving one biological macromolecule fragment of the target biological macromolecule immobilized to the solid support through the chemically cleavable linker and a second biological macromolecule fragment, from the opposite end of the target biological macromolecule, immobilized through the photocleavable linker. Each fragment then can be further manipulated using a process as disclosed herein or can be analyzed by IR-MALDI mass spectrometry following sequential cleavage of the fragments, for example, after first cleaving the chemically cleavable linker, then cleaving the photocleavable linker. Such a process provides a convenient means of analyzing both termini of a biological macromolecule, thereby facilitating analysis of the target biological macromolecule.

Immobilization of a target biological macromolecule at both termini can be performed by modifying both ends of the biological macromolecule, for example, one terminus being modified to allow formation of a chemically cleavable linkage with the solid support and the other terminus being modified to allow formation of a photocleavable linkage with the solid support. Alternatively, the biological macromolecules can be split into two portions, one portion being modified at one terminus allow formation, for example, of a chemically cleavable linkage, and the second portion being modified at the other terminus to allow formation, for example, of a photocleavable linkage. The two populations of modified biological macromolecules then can be immobilized, together, on a solid support containing the appropriate functional groups for completing immobilization.

IR-MALDI MASS SPECTROMETRIC ANALYSIS OF BIOLOGICAL MACROMOLECULES

The processes disclosed herein are useful for analyzing a biological macromolecule by subjecting a composition containing the biological macromolecule and a liquid matrix, which absorbs IR radiation, to IR-MALDI mass spectrometry. Depending on the process selected, the presence of a biological macromolecule can be detected, for example, in a biological sample; or a particular biological macromolecule can be identified, for example, by comparison to a corresponding known biological macromolecule, or by determining its molecular mass or at least a part of its subunit sequence (see, for example, U.S. Patent Nos. 5,503,980; 5,547,835; 5,605,798; and 5,691,194; see, also, International Publs. WO 94/16101; WO 94/21822, WO 96/29431; WO 97/37041; WO 97/42348; and WO 98/20019).

Mass spectrometric analysis using an IR laser

The support containing a sample can be placed in a vacuum chamber of a mass analyzer to identify or detect the nucleic acid in the sample. Preferably, the mass analyzer can maintain the temperature of a sample at a preselected value, for example, a temperature in the range of at least about -200°C to about 80°C, preferably at least about -60° C to about 40° C, more preferably -200° C to about 20° C, and most preferably about -60° C to about 20° C, during

sample preparation, disposition and/or analysis. For example, improved spectra may be obtained, in some instances, by cooling the sample to a temperature below room temperature during sample preparation or mass analysis. Further, as described above, the vacuum stability of a matrix may be increased by cooling. Alternatively, it may be useful to heat a sample to denature double stranded nucleic acids into single strands or to decrease the viscosity during sample preparation.

Desorption and ionization of the sample is performed in the mass analyzer using infrared radiation. Preferred infrared wavelengths include in the are in the mid-IR wavelength region, from about 2.5 μm to about 12 μm. Preferred sources of infrared radiation are CO lasers, which emit at about 6 μm; CO_2 lasers, which emit at about 9.2 μm to 11 μm; Er lasers, with any of a variety of crystals, for example, Er-YAG (yttrium-aluminum-garnet), Er-YILF or Er-YSGG, emitting at wavelengths about 3 μm; and optical paramagnetic oscillator lasers emitting in the range of about 2.5 μm to about 12 μm.

Pulse duration, field strength and other parameters

Solid state Erbium lasers with pulse widths around 100 ns can be used for infrared Matrix-Assisted Laser Desorption/Tonization mass spectrometry (IR-MALDI MS) [Overberg et al, *Rapid Commun. Mass Spectrom.*, 1990, 4, 293-296; Berkenkamp et al., *Rapid Commun. Mass Spectrom.*, **1997**, 11, 1399-1406]. Optical parametric oscillators (OPO) with pulse durations of a few nanoseconds may also be used in IR-MALDI MS. The fixed pulse width of the OPO systems of a few nanoseconds is determined by the pump laser. The pulse duration and/or size of the irradiated area (spot size) can be varied to generate multiple charged ions. A preferred pulse duration is in the range of about 100 picoseconds (psec) to about 500 nanoseconds (ns).

An Er:YAG- and an OPO laser were used to investigate pulse width and wavelength dependence of IR-MALDI-MS in the 5-200 ns pulse width and 3 μm wavelength region. For laser pulse durations from 90 to 185 ns an Er:YAG laser (Spektrum GmbH, Berlin, Germany, wavelength $\lambda = 2.94$ μm) was used. The pulse duration was varied by changing the Q-switch delay time. For the Nd:YAG pumped OPO laser (Mirage 3000B, Continuum, Santa Clara, CA, USA)

the pulse width was fixed at 6 ns, whereas this system is tunable from 2.2 μm to 4.0 μm. The wavelength scale was calibrated to an accuracy of ± 5 nanometers. An in-house-built TOF instrument with a linear (2.2 m) and a reflectron port (3.5 m equivalent flight length) was used. The mass spectrometer can be operated with static or delayed ion extraction. Special optics were implemented to permit a rapid interchange of the two laser beams. A 150 μm pinhole was illuminated by the central part of the Gaussian beams and imaged onto the sample to ensure a homogeneous and equal sample illumination for both lasers. All spectra were obtained under identical instrumental conditions and from identical samples.

Results: a) To a first approximation the threshold fluences for the generation of Cytochrome C mass spectra were independent of the pulse duration in the range of 6 to 185 ns.

Laser System	H_0/J m^{-2}		
	Succinic acid	Thiourea	Glycerol
OPO (τ = 6 ns)	3564 ± 695	2053 ± 296	4186 ± 143
Er:YAG (τ = 98 ns)	4304 ± 538	3433 ± 127	4992 ± 118
Er:YAG (τ = 185 ns)	4591 ± 532	3398 ± 398	4941 ± 730

For the OPO-systems the threshold fluences were consistently and statistically significantly lower by up to a factor of 1.5 as compared to the Er:YAG laser. However, the irradiances of ~50 MW/cm^2 (τ = 6 ns) for the OPO system and of ~2 MW/cm^2 (τ = 18.5 ns) for the Er:YAG laser differ by a factor of ~25. It is, therefore, concluded, that the desorption in IR-MALDI is governed by the deposited energy per unit volume, rather than the peak power or irradiance for pulse durations up to 200 ns.

b) Within the experimental error, mass resolution for signals of peptides, desorbed out of a succinic acid matrix, was observed to be independent of the pulse width within the range of 6 - 100 ns for static and delayed ion extraction. For longer pulses up to 200 ns and static ion extraction the resolution decreased by up to a factor of two. In the analysis of the

influence of laser pulse widths on the peak resolution of Gramicidin S, an optimal resolution of $m/\Delta m$ = 11000 was observed for 6 ns OPO laser pulses with delayed ion extraction, as well as for 100 ns Erbium laser pulses in the linear mode of the mass spectrometer.

c) For the 6 ns pulses an increase in the abundance of multiply charged ions and a decrease of signals of oligomers was observed, as compared to 100 ns pulses.

d) The threshold fluence for the generation of IR-MALDI spectra was determined in the wavelength range from 2.6 μm to 3.6 μm for several solid and liquid matrices with the OPO laser system. They were compared to the corresponding transmission spectra of the matrices [Merke, R., Langenbucher, F., *Infrared Spectra*, Heyden & Co., Freiburg, 1964]. A clear correlation between the threshold fluences for succinic acid and glycerol on their (inverse) transmission was observed in a study of the influence of laser wavelength λ on the threshold fluence H_0 of cytochrome C. For glycerol the double peak structure is clearly reproduced. A similar behavior was observed for triethanolamine. For succinic acid the threshold fluence follows the absorption spectrum in the range of 3.2 - 3.6 μm. The surprisingly low threshold fluence between 2.8 and 3.2 μm seems to reflect the strong absorption of residual water in the succinic acid microcrystals.

Field strengths typically less than 1000 V/mm, preferably as low as 200 V/mm, particularly for proteins, are used.

A preferred spot size is in the range of about 50 μm in diameter to about 1 mm. IR-MALDI can be matched with an appropriate mass analyzer, including linear (lin) or reflector (ref), with linear and nonlinear fields, for example, curved field reflectron, time-of-flight (TOF), single or multiple quadrupole, single or multiple magnetic sector, Fourier transform ion cyclotron resonance or ion trap. Preferably, detection is performed using a linTOF or a refTOF mode instrument in positive or negative ion modes, so that the ions are accelerated through a total potential difference of about 3 kV to about 30 kV in the split extraction source using static or delayed ion extraction (DE). TOF mass spectrometers separate ions according to their mass-to-charge ratio by measuring the time it

takes generated ions to travel to a detector. The technology behind TOF mass spectrometers is described for example in U.S. Patent Nos. 5,627,369; 5,625,184; 5,498,545; 5,160,840 and 5,045,694. Delayed extraction with delay time ranging from about 50 nsec to about 5 μsec may improve the mass resolution of some nucleic acids, for example, nucleic acids in the mass range of from about 30 kDa to about 50 kDa, using either a liquid or solid matrix. For delayed extraction, conditions are selected to permit a longer optimum extraction delay and hence a longer residence time, which results in increased resolution (see, e.g., Juhasz et al., Anal. Chem. 68:941-946 (1996); Vestal et al., Rapid Commun. Mass Spectrom. 9:1044-1050 (1995); see, also, U.S. Patent Nos. 5,777,325; 5,742,049; 5,654,545; 5,641,959; 5,654,545; and 5,760,393, for descriptions of MALDI and delayed extraction protocols). In delayed ion extraction, a time delay is introduced between the formation of the ions and the application of the accelerating field. During the time lag, the ions move to new positions according to their initial velocities. By properly choosing the delay time and the electric fields in the acceleration region, the time of flight of the ions can be adjusted so as to render the flight time independent of the initial velocity to the first order.

ANALYSIS OF NUCLEIC ACIDS BY IR-MALDI

Methods and processes for sequencing, diagnosis and detection of nucleic acids using UV MALDI have been developed and are known to those of skill in the art (see, *e.g.*, U.S. Patent Nos. 5,605,798, 5,830,655, 5,700,642, allowed U.S. application Serial No. 08/617,256, published International PCT application Nos. WO 96/29431, WO 98/20019, WO 99/14375, WO 97/03499, WO 98/26095 and others).

Processes of using IR-MALDI to analyze a nucleic acid in a liquid matrix are provided. Nucleic acids to be analyzed according to a process provided herein can include any single stranded or double stranded polynucleotide such as DNA, including genomic DNA and cDNA; RNA; or an analog of RNA or DNA, as well as nucleotides or nucleosides and any derivative thereof. Nucleic acids can be of any size ranging from single nucleotides or nucleosides to tens of

thousands of base pairs. For analysis herein, preferred nucleic acids contain about one thousand nucleotides or less.

Nucleic acids may be obtained from a biological sample, which can be any material obtained from any living source, including a human, animal, plant, bacterium, fungus, protist or virus, using any of a number of procedures that are well known in the art. A particular isolation procedure for obtaining a nucleic acid from a biological sample can be selected as appropriate for the particular biological sample. For example, freeze-thaw or alkaline lysis procedures can be useful for obtaining nucleic acid molecules from solid materials; heat and alkaline lysis procedures can be useful for obtaining nucleic acids from blood (Rolff et al., *PCR: Clinical Diagnostic and Research* (Springer Verlag 1994)).

Prior to being mixed with a liquid matrix, the particular nucleic acid to be analyzed may be further processed to yield a relatively pure, isolated nucleic acid sample. For example, a standard ethanol precipitation may be performed on restriction enzyme digested DNA. Alteratively, PCR products may require primer removal prior to analysis. Likewise, RNA strands can be separated from the molar excess of premature termination products always present in *in vitro* transcription reactions.

SEQUENCING

Exemplary formats and strategies

Any sequencing strategy known to those of skill in the art, including Sanger, exonuclease and hybridization methods can be adapted for use with IR MALDI methods provided herein, by liquid matrices and and IR MALDI. For example, a Sanger sequencing strategy assembles the sequence information by analysis of the nested fragments obtained by base-specific chain termination via their different molecular masses, which can be determined using IR-MALDI. Further increases in throughput, if needed can be obtained by conditioning the nucleic acid fragments, such as by introducing mass modifications into the oligonucleotide primer, the chain-terminating nucleoside triphosphates and/or the chain-elongating nucleoside triphosphates, as
well as using integrated tag sequences that allow multiplexing by

hybridization of tag specific probes with mass differentiated molecular weights.

Exonuclease-based sequencing protocols can also be performed. These methods, which include those described in U.S. Patent No. 5,622,824 adapted for use with IR-MALDI, involve a direct sequencing approach and can begin wit DNA fragments cloned into conventional cloning vectors. The DNA is by means of protection, specificity of enzymatic activity, or immobilization, unilaterally degraded in a stepwise manner via exonuclease digestion and the nucleotides or derivatives detected by mass spectrometry. Prior to the enzymatic degradation, sets of ordered deletions that span the whole sequence of the cloned DNA fragment are created. In this manner, mass-modified nucleotides can be incorporated using a combination of exonuclease and DNA/RNA polymerase. This permits either multiplex mass spectrometric detection, or modulation of the activity of the exonuclease so as to synchronize the degradative process.

Methods for sequencing by hybridization include methods of positional sequencing by hybridization (see, *e.g.*, U.S. Patent No. 5,503,980, 5,795,714 and 5,631,134). Briefly, sequencing by hybridization refers to methods methods of sequencing a nucleic acid by hybridizing that nucleic acid with a set of nucleic acid probes containing random, but determinable sequences within the single stranded portion adjacent to a double stranded portion where the single stranded portion of the set preferably comprises every possible combination of sequences over a predetermined range. Hybridization occurs by complementary recognition of the single stranded portion of a target with the single stranded portion of the probe and is thermodynamically favored by the presence of adjacent double strandedness of the probe. In particular, a method for determining a nucleotide sequence of a nucleic acid target by hybridization includes the steps of creating a set of nucleic acid probes, wherein each probe is preferably about 14-50 nucleotides in length and has a double stranded portion, a single stranded portion, and a variable sequence within the single stranded portion that

is determinable; hybridizing the target that is at least partly single stranded to one or more of the nucleic acid probes; and determining the nucleotide sequence of the target that is hybridized to the single stranded portion of any probe. To detect the probes the target can be labeled with a first detectable label at a terminal site and a second different detectable label at an internal site. The labels are selected to be detectable by IR mass spectrometry.7

Examples of the above formats

In one exemplary direct sequencing embodiment, the method of sequencing obtaining multiple nucleic acid copies of the target nucleic acid, where the multiple copies contain at least one mass modified nucleotide, corresponding to one of the four possible nucleotide bases; cleaving the multiple nucleic acid copies from a first end to a second end with an exonuclease having an activity, which is inhibited by the mass-modified nucleotide, thereby generating base terminated nucleic acid fragments; identifying the nested nucleic acid fragments by IR-MALDI; and (iv) determining the sequence of the target nucleic acid from the identified nested nucleic acid fragments.

In all formats, the nucleic acids can be immobilized, including in array formats. Immobilization can be effected with linkers that are cleavable, such as by the IR radiation emitted by the IR laser. The linkages can be reversible or irreversible.

Thus, processes for determining a subunit sequence of a target biological macromolecule also are provided. A sequence of a target biological macromolecule can be determined by contacting the biological macromolecule with an agent that cleaves the biological macromolecule unilaterally from a terminus of the biological macromolecule, to produce a nested set of deletion fragments; preparing a composition containing the nested set of biological macromolecule fragments and a liquid matrix, which absorbs infrared radiation; determining the molecular weight value of each biological macromolecule fragment in the composition by IR-MALDI mass spectrometry; and determining

the sequence of the nucleic acid from the molecular weight values of the biological macromolecule fragments in the set.

A sequence of a target nucleic acid, for example, can be determined by subjecting the target nucleic acid to exonuclease digestion for various periods of time to produce a nested set of deletion fragments containing the target nucleic acid sequence (see International Publ. WO 94/21822), then analyzing the nested set of deletion fragments by IR-MALDI. Similarly, a sequence of a target polypeptide can be determined by subjecting the polypeptide to an exopeptidase, which can be a carboxypeptidase such as carboxypeptidase Y, carboxypeptidase P, carboxypeptidase A, carboxypeptidase G or carboxypeptidase B; or an aminopeptidase such as alanine aminopeptidase, leucine aminopeptidase, pyroglutamate peptidase, dipeptidyl peptidase and microsomal peptidase; or a chemical polypeptide fragmenting agent such as phenylisothiocyanate, for various periods of time to produce a nested set of fragments of the biological macromolecule, which can be analyzed by IR-MALDI mass spectrometry to determine the sequence of the target biological macromolecule (see, also, *Protein LabFax*, pages 273-276 (ed., N.C. Price; Bios Scientific Publ., 1996); listing polypeptide fragmenting agents). Exonucleases, exopeptidases and exoglycosidases are well known in the art (see, for example, U.S. Patent No. 5,821,063), as are methods of modifying the activity of such agents (see, for example, U.S. Patent No. 5,792,664; International Publ. WO 96/36732).

A sequence of a target biological macromolecule also can be determined by treating the biological macromolecule with an agent that cleaves the biological macromolecule unilaterally from a terminus, in a time-limited manner, and identifying the released monomer subunits by IR-MALDI mass spectrometry. If desired, degradation of a target biological macromolecule can be performed in a reactor apparatus (see International Publ. WO 94/21822), in which the biological macromolecule can be free in composition and the agent that cleaves can be immobilized, or in which the agent that cleaves can be free in composition and the biological macromolecule can be immobilized. At time intervals or as a continuous stream, the reaction mixture containing released

subunits is transported from the reactor for analysis by IR-MALDI mass spectrometry. Prior to IR-MALDI mass spectrometric analysis, the released subunits can be transported to a reaction vessel for conditioning, which can be by mass modification.

A sequence of a target biological macromolecule also can be determined by generating at least two biological macromolecule fragments from the target biological macromolecule; preparing a composition containing the biological macromolecule fragments and a liquid matrix, which absorbs infrared radiation; and analyzing the biological macromolecule fragments in the composition by IR-MALDI mass spectrometry, thereby determining the sequence of the target nucleic acid molecule. In particular, such a process can be useful for determining the order of subunit sequences within a large biological macromolecule sequence (see International Publ. WO 98/20019).

A process of determining the subunit sequence of at least one species of target biological macromolecule, i, also is provided. Such a process can be performed, for example, by contacting the species of target biological macromolecule with one or more agents sufficient to cleave each the bonds between each monomer subunit in the target biological macromolecule, to produce a nested set of deletion fragments; preparing a composition containing at least one biological macromolecule fragment of the set and a liquid matrix, which absorbs infrared radiation; and determining the molecular mass of the at least one biological macromolecule fragment by IR-MALDI mass spectrometry; and repeating these steps until the molecular mass of each biological macromolecule fragment in said set has been determined, thereby determining the subunit sequence of the species of target biological macromolecule. Such a process is particularly suitable for multiplex analysis of a plurality of $i+1$ species of target biological macromolecules. For multiplex analysis, each species of target biological macromolecule can be differentially mass modified such that a biological macromolecule fragment of each species of target biological macromolecule can be distinguished from every other biological macromolecule species by IR-MALDI mass spectrometry.

A process of determining the nucleotide sequence of at least one species of nucleic acid also is provided. Such a process can be performed by synthesizing complementary nucleic acids, which are complementary to the species of nucleic acid to be sequenced, starting from an oligonucleotide primer and in the presence of chain terminating nucleoside triphosphates, to produce four sets of base-specifically terminated complementary polynucleotide fragments; preparing a composition for IR-MALDI that contains four sets of polynucleotide fragments and a liquid matrix, which absorbs infrared radiation; determining the molecular weight value of each polynucleotide fragment by IR-MALDI mass spectrometry; and determining the nucleotide sequence of the species of nucleic acid by aligning the molecular weight values according to molecular weight. The process is particularly suitable to multiplex analysis of a plurality of i + 1 species of nucleic acids, which can be sequenced concurrently using i + 1 primers. For multiplex analysis, one of the i + 1 primers is an unmodified primer or a mass modified primer, and the other i primers are mass modified primers, such that each of the i + 1 primers can be distinguished from every other primer by IR-MALDI mass spectrometry.

A sequence of a target nucleic acid also can be determined by hybridizing at least one partially single stranded target nucleic acid to one or more nucleic acid probes, each probe containing a double stranded portion, a single stranded portion, and a determinable variable sequence within the single stranded portion, to produce at least one hybridized target nucleic acid; preparing a composition containing the hybridized target nucleic acid and a liquid matrix, which absorbs infrared radiation; and determining a sequence of the hybridized target nucleic acid by IR-MALDI mass spectrometry based on the determinable variable sequence of the probe to which the target nucleic acid hybridized (U.S. Patent No. 5,503, 980). Optionally, a hybridized target nucleic acid can be ligated to the determinable variable sequence. If desired, the steps of the process can be repeated a sufficient number of times to determine an entire sequence of a target nucleic acid. Where a plurality of target nucleic acids are to be sequenced, the one or more nucleic acid probes can be immobilized in an array.

IR-MALDI mass spectrometry also can be used to determine a nucleic acid sequence by analyzing a target polypeptide encoded by the nucleic acid. Since the mass of a polypeptide is only about 10% of the mass of its encoding nucleic acid, the translated polypeptide can be more amenable to mass spectrometric detection. In addition, IR-MALDI mass spectrometric detection of polypeptides can yield analytical signals of high sensitivity and resolution (see Berkenkamp et al., Rapid Commun. Mass Spectrom. 11:1399-1406 (1997)).

Oligonucleotide sizing, fingerprinting and sequencing using IR-MALDI mass spectrometry and immobilized cleavable primers

IR-MALDI mass spectrometry can also be used, in conjunction with the immobilized cleavable primers described in U.S. Patent No. 5,830,655 and U.S. Patent No. 5,700,642 or other such primers, to determine the size of a primer extension product. In one specific embodiment, a method for determining the size of a primer extension product is provided. It includes the steps of (a) hybridizing a primer with a target nucleic acid, where the primer (i) is complementary to the target nucleic acid; (ii) has a first region containing the 5′ end of the primer, and (iii) has a second region containing the 3′ end of the primer, where the 3′ end is capable of serving as a priming site for enzymatic extension and where the second region contains a selected cleavable site; (b) extending the primer enzymatically to generate a polynucleotide mixture containing an extension product composed of the primer and an extension segment; (c) cleaving the extension product at the cleavable site to release the extension segment; and (d) sizing the extension segment by IR-MALDI mass spectrometry with a liquid matrix, whereby the cleaving is effective to increase the read length of the extension segment relative to the read length of the product of (b).

In one embodiment, the target nucleic acid contains an immobilization attachment site and is thereby immobilized by attachment to a solid support. The target nucleic acid can be immobilized prior to the extending. Also preferably, the target nucleic acid is immobilized prior to the cleaving. Further more preferably, the product of (b) from the immobilized target nucleic acid is separated prior to the cleaving step.

In another embodiment, the cleavable site is a nucleotide capable of blocking 5′ to 3′ enzyme-promoted digestion, and where the cleaving is carried out by digesting the first region of the primer with an enzyme having a 5′ to 3′ exonuclease activity. In another embodiment, the cleavable site is located at or within about five nucleotides from the 3′ end of the primer. More preferably, the second region of the primer is a single nucleotide that also contains the cleavable site, such as, but are not limited to, a ribonucleotide, dialkoxysilane, 3′-(S)-phosphorothioate, 5′-(S)-phosphorothioate, 3′-(N)-phosphoramidate, 5′-(N)phosphoramidate, uracil or ribose. The enzyme for extending the primer in step (b) can be a DNA polymerase.

In yet another embodiment, the extending is carried out in the presence of a nucleotide containing (i) an immobilization attachment site and (ii) a releasable site, which is thereby incorporated into the extension segment. More preferably, a further step of immobilizing the extension segment at the immobilization attachment site and releasing the extension segment at the releasable site prior to the sizing by IR-MALDI mass spectrometry is included.

In another specific embodiment, a method for determining the size of a primer extension product is provided, which method comprises (a) hybridizing a primer with a target nucleic acid, where the primer (i) is complementary to the target nucleic acid; (ii) has a first region containing the 5′ end of the primer, and an immobilization attachment site, where the immobilization attachment site of the primer is composed of a series of bases complementary to an intermediary oligonucleotide, and (iii) has a second region containing the 3′ end of the primer, where the 3′ end is capable of serving as a priming site for enzymatic extension and where the second region contains a selected cleavable site, (b) extending the primer enzymatically to generate a polynucleotide mixture containing an extension product composed of the primer and an extension segment; (c) cleaving the extension product at the cleavable site to release the extension segment, where prior to the cleaving the primer is immobilized by specific hybridization of the immobilization attachment site to the intermediary oligonucleotide bound to a solid support; and (d) sizing the extension segment by IR-MALDI mass spectrometry with a liquid matrix, whereby the cleaving is

effective to increase the read length of the extension segment relative to the read length of the product of (b).

In still another specific embodiment, a method for determining the size of a primer extension product is provided that includes (a) combining first and second primers with a target nucleic acid, under conditions that promote hybridization of the primers to the nucleic acid, generating primer/nucleic acid complexes, where the first primer (i) has a 5′ end and a 3′ end, (ii) is complementary to the target nucleic acid, (iii) has a first region containing the 5′ end of the first primer and (iv) has a second region containing the 3′ end of the first primer, where the 3′ end is capable of serving as a priming site for enzymatic extension and where the second region contains a cleavable site, and where the second primer (i) has a 5′ end and a 3′ end, (ii) is homologous to the target nucleic acid, (iii) has a first segment containing the 3′ end of the second primer, and (iv) has a second segment containing the 5′ end of the second primer and an immobilization attachment site; (b) converting the primer/nucleic acid complexes to double-stranded fragments in the presence of a DNA polymerase and deoxynucleoside triphosphates; (c) amplifying the primer-containing fragments by successively repeating the steps of (i) denaturing the double-stranded fragments to produce single-stranded fragments, (ii) hybridizing the single stranded fragments with the first and second primers to form strand/primer complexes, (iii) generating amplification products from the strand/primer complexes in the presence of DNA polymerase an deoxynucleoside triphosphates, and (iv) repeating steps (i) to (iii) until a desired degree of amplification has been achieved; (d) immobilizing amplification products containing the second primer via the immobilization attachment site; (e) removing non-immobilized amplified fragments; (f) cleaving the immobilized amplification products at the cleavable site, to generate a mixture including a double-stranded product; (g) denaturing the double-stranded product to release the extension segment; and (h) sizing the extension segment by IR-MALDI mass spectrometry with a liquid matrix, whereby the cleaving is effective to increase the read length of the extension segment relative to the read length of the amplified strand-primer complexes of (c).

In another embodiment, the method for determining the size of a includes the steps of (a) hybridizing a primer with a target nucleic acid, where the primer (i) is complementary to the target nucleic acid; (ii) has a first region containing the 5′ end of the primer and an immobilization attachment site, and (iii) has a second region containing the 3′ end of the primer, where the 3′ end is capable of serving as a priming site for enzymatic extension and where the second region contains a selected cleavable site, (b) extending the primer enzymatically to generate a polynucleotide mixture containing an extension product composed of the primer and an extension segment; (c) cleaving the extension product at the cleavable site to release the extension segment, where prior to the cleaving the primer is immobilized at the immobilization attachment site; and (d) sizing the extension segment by IR-MALDI mass spectrometry with a liquid matrix, whereby the cleaving is effective to increase the read length of the extension segment relative to the read length of the product of (b). The enzyme for extending the primer in step (b) can be a DNA polymerase.

In one embodiment, the cleavable site is located at or within about five nucleotides from the 3′ end of the primer. More preferably, the second region of the primer is a single nucleotide that also contains the cleavable site, such as, but are not limited to, a ribonucleotide, dialkoxysilane, 3′-(S)-phosphorothioate, 5′-(S)phosphorothioate, 3′-(N)-phosphoramidate, 5′-(N)phosphoramidate, or ribose.

In another embodiment, a further step of washing the immobilized product prior to the cleaving step is included. In another embodiment, the primer is immobilized on a solid support by attachment at the immobilization attachment site to an intervening spacer arm bound to the solid support. More preferably, the intervening spacer arm is six or more atoms in length. The immobilization attachment site preferably occurs as a substituent on one of the bases or sugars of the DNA primer. In another embodiment, the immobilization attachment site is biotin or digoxigenin. In another embodiment, the primer is immobilized on a solid support, including, but are not limited to, glass, silicon, polystyrene, aluminum, steel, iron, copper, nickel or gold.

In another embodiment, the method for determining the size of a primer includes the steps of: (a) combining first and second primers with a target nucleic acid under conditions that promote the hybridization of the primers to the nucleic acid, thus generating primer/nucleic acid complexes, where the first primer (i) is complementary to the target nucleic acid; (ii) has a first region containing the 5′ end of the primer and an immobilization attachment site, and (iii) has a second region containing the 3′ end of the primer, where the 3′ end is capable of serving as a priming site for enzymatic extension and where the second region contains a cleavable site, and where the second primer is homologous to the target nucleic acid; (b) converting the primer/nucleic acid complexes to double-stranded fragments in the presence of a suitable polymerase and all four dNTPs; (c) amplifying the primer-containing fragments by successively repeating the steps of (i) denaturing the double-stranded fragments to produce single-strand fragments, (ii) hybridizing the single strands with the primers to form strand/primer complexes, (iii) generating double-stranded fragments from the strand/primer complexes in the presence of DNA polymerase and all four dNTPs, and (iv) repeating steps (i) to (iii) until a desired degree of amplification has been achieved; (d) denaturing the amplified fragments to generate a mixture including a product composed of the first primer and an extension segment; (e) immobilizing amplified fragments containing the first primer, utilizing the immobilization attachment site, and removing non-immobilized amplified fragments; (f) cleaving the immobilized fragments at the cleavable site to release the extension segment; and (g) sizing the extension segment by IR-MALDI mass spectrometry with a liquid matrix, whereby the cleaving is effective to increase the read length of the extension segment relative to the read length of the product of (d).

In another embodiment, a method for determining a single base fingerprint of a target DNA sequence is provided. The method includes the steps of (a) hybridizing a primer with a target DNA, where the primer (i) is complementary to the target DNA; (ii) has a first region containing the 5′ end of the primer and an immobilization attachment site, and (iii) has a second region containing the 3′ end of the primer, where the 3′ end is capable of serving as a

priming site for enzymatic extension and where the second region contains a selected cleavable site; (b) extending the primer with an enzyme in the presence of a dideoxynucleoside triphosphate corresponding to the single base, to generate a polynucleotide mixture of primer extension products, each product containing a primer and an extension segment; (c) cleaving the extension products at the cleavable site to release the extension segments, where prior to the cleaving the primers are immobilized at the immobilization attachment sites; (d) sizing the extension segments by IR-MALDI mass spectrometry with a liquid matrix, whereby the cleaving is effective to increase the read length of any given extension segment relative to the read length of its corresponding primer extension product of (b), and (e) determining the positions of the single base in the target DNA by comparison of the sizes of the extension segments.

In another embodiment, a method for an adenine fingerprint of a target DNA sequence by (a) hybridizing a primer with a DNA target, where the primer (i) is complementary to the target DNA; (ii) has a first region containing the 5′ end of the primer and an immobilization attachment site, and (iii) has a second region containing the 3′ end of the primer, where the 3′ end is capable of serving as a priming site for enzymatic extension and where the second region contains a selected cleavable site; (b) extending the primer with an enzyme in the presence of deoxyadenosine triphosphate (dATP), deoxythymidine triphosphate (dTTP), deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP), and deoxyuridine triphosphate (dUTP), to generate a polynucleotide mixture of primer extension products containing dUTP at positions corresponding to dATP in the target, each product containing a primer and an extension segment; (c) treating the primer extension products with uracil DNA-glycosylase to fragment specifically at dUTP positions to produce a set of primer extension degradation products; (d) washing the primer extension degradation products, where prior to the washing, the primer extension degradation products are immobilized at the immobilization attachment sites, each immobilized primer extension degradation product containing a primer and an extension segment, where the washing is effective to remove non-immobilized species; (e) cleaving the immobilized primer extension

degradation products at the cleavable site to release the extension segments; (f) sizing the extension segments by IR-MALDI mass spectrometry with a liquid matrix, whereby the cleaving is effective to increase the read length of any given extension segment relative to the read length of its corresponding primer extension degradation product; and (g) determining the positions of adenine in the target DNA by comparison of the sizes of the released extension segments.

In another specific embodiment, a method for determining the DNA sequence of a target DNA sequence is provided, which method comprises (a) hybridizing a primer with a target DNA, where the primer (i) is complementary to the target DNA; (ii) has a first region containing the 5′ end of the primer and an immobilization attachment site, and (iii) has a second region containing the 3′ end of the primer, where the 3′ end is capable of serving as a priming site for enzymatic extension and where the second region contains a cleavable site, (b) extending the primer with an enzyme in the presence of a first of four different dideoxy nucleotides to generate a mixture of primer extension products each product containing a primer and an extension segment; (c) cleaving at the cleavable site to release the extension segments, where prior to the cleaving the primers are immobilized at the immobilization attachment sites; (d) sizing the extension segments by IR-MALDI mass spectrometry with a liquid matrix, whereby the cleaving is effective to increase the read length of the extension segment relative to the read length of the product of (b), (e) repeating steps (a) through (d) with a second, third, and fourth of the four different dideoxy nucleotides, and (f) determining the DNA sequence of the target DNA by comparison of the sizes of the extension segments obtained from each of the four extension reactions.

In yet another specific embodiment, a method for determining the DNA sequence of a target DNA sequence is provided, which method comprises (a) hybridizing a primer with a target DNA, where the primer (i) is complementary to the target DNA; (ii) has a first region containing the 5′ end of the primer and an immobilization attachment site, and (iii) has a second region containing the 3′ end of the primer, where the 3′ end is capable of serving as a priming site for enzymatic extension and where the second region contains a cleavable site, (b)

extending the primer with an enzyme in the presence of a first of four different deoxynucleoside α-thiotriphosphate analogs (dNTPαS) to generate a mixture of primer extension products containing phosphorothioate linkages, (c) treating the primer extension products with a reagent that cleaves specifically at the phosphorothioate linkages, where the treating is carried out under conditions producing limited cleavage, resulting in the production of a group of primer extension degradation products, (d) washing the primer extension degradation products, where prior to the washing, the primer extension degradation products are immobilized at the immobilization attachment sites, each immobilized primer extension degradation product containing a primer and an extension segment, where the washing is effective to remove non-immobilized species, (e) cleaving at the cleavable site to release the extension segments, (f) sizing the extension segments by IR-MALDI mass spectrometry with a liquid matrix, whereby the cleaving is effective to increase the read length of any given extension segment relative to the read length of its corresponding primer extension degradation product, (g) repeating steps (a) through (f) with a second, third, and fourth of the four different dNTPαSs, and (h) determining the DNA sequence of the target DNA by comparison of the sizes of the extension segments obtained from each of the four extension reactions. More preferably, the reagent of step (c) is exonuclease, 2-iodoethanol, or 2,3-epoxy-1-propanol.

DIAGNOSIS AND DETECTION

Diagnostics

Using a process as disclosed herein, accurate (at least about 1% accurate) masses of a DNA sample can be obtained for at least about 2000-mer DNA (masses of at least about 650 kDa) and at least about 1200-mer RNA (masses of at least about 400 kDa; see Example 1). In addition, signals of single stranded, as well as double stranded, nucleic acids can be obtained in the spectra (see Figure 3). The improved accuracy for measuring the mass of DNA by IR-MALDI mass spectrometry (accuracy of at least about 1%) is far superior to that provided by standard agarose gel sizing of nucleic acids (accuracy of about 5%). The accuracy of mass determination of RNA by IR-MALDI mass spectrometry (accuracy of at least about 0.5%) is even more significant, since

an accurate size determination of RNA by gel analysis is difficult, if not impossible, in part because of the absence of suitable size markers and of a sufficiently suitable gel matrix.

In addition to the extension in mass range obtained using a process as disclosed herein, there is a dramatic decrease in the amount of analyte needed for preparation of the sample for mass spectrometry, down to the low femtomole (fmol) or attomole (attomol) range, even with an essentially simple preparation method. Also, by using a liquid matrix rather than a solid matrix, the ion signals generated are more reproducible from shot to shot. Use of a liquid matrix also facilitates sample dispensation, for example, onto various fields of a chip array. Furthermore, by using a liquid matrix in conjunction with IR-MALDI mass spectrometry, essentially all sample left on the target after IR-MALDI analysis can be retrieved for further use.

DIAGNOSIS AND DETECTION

A process of determining the molecular mass of a target biological macromolecule by IR-MALDI mass spectrometry is provided. Such a process can be performed, for example, by preparing a composition for IR-MALDI containing the biological macromolecule to be analyzed and a liquid matrix, which absorbs infrared radiation; and analyzing the biological macromolecule in the composition by IR-MALDI mass spectrometry (see Example 1; see, also, Berkenkamp et al., Rapid Commun. Mass Spectrom. 11:1399-1406 (1997); Berkenkamp et al., Science 281:260-262 (1998)). The molecular mass of the target biological macromolecule is determined by running, in parallel or in a separate spectrum, one or more control biological macromolecules having known molecular masses, and comparing the spectrum produced by the target spectrum with the spectrum of the control biological macromolecules. A control biological macromolecule, which can be a corresponding known biological macromolecule, generally is of the same type of molecule as the target biological macromolecule, for example, each is a nucleic acid or each is polypeptide. The control biological macromolecule need not be the same type of molecule as a target biological macromolecule in order to determine the molecular mass of the target biological macromolecule (see Example 1).

IR-MALDI mass spectrometry also can be used for detecting a target biological macromolecule by preparing a composition containing a biological macromolecule and a liquid matrix, which absorbs infrared radiation; and performing IR-MALDI mass spectrometry on the composition to identify the target biological macromolecule in the composition, thereby detecting the target biological macromolecule. If desired, the target biological macromolecule can be present in or isolated from a biological sample. Accordingly, a process for identifying the presence of a target biological macromolecule in a biological sample also is provided.

The presence of a target biological macromolecule, for example, a nucleic acid in a biological sample can be identified by preparing a composition for IR-MALDI, containing a biological sample containing nucleic acid molecules (or nucleic acid molecules isolated from the biological sample) and a liquid matrix, which absorbs infrared radiation; then analyzing the composition by IR-MALDI mass spectrometry. Detection of a nucleic acid molecule having a molecular mass of the target nucleic acid sequence identifies the presence of the target nucleic acid sequence in the biological sample. The molecular mass of the target biological macromolecule can be determined by comparison to a control spectrum, or can be determined based on the spectrum produced by a corresponding known biological macromolecule. Alternatively, a sequence of the biological macromolecule can be determined, thereby identifying the presence of the biological macromolecule.

Since the processes disclosed herein allow a characterization of a target biological macromolecule obtained from a biological sample, IR-MALDI mass spectrometry can be used to identify an individual having a disease or condition, or a predisposition to a disease or condition, by detecting a characteristic of a target biological macromolecule that is associated with the disease or the condition. Such a process can be performed, for example, by preparing a composition for IR-MALDI, containing the biological macromolecule, which is obtained from an individual to be tested, and a liquid matrix, which absorbs infrared radiation; and analyzing the biological macromolecule, or a relevant portion of the biological macromolecule, in the composition by IR-MALDI mass

spectrometry. A determination of a particular mass of the target biological macromolecule identifies the individual as having the disease or condition or a predisposition to the disease or condition. Such a process is particularly useful for identifying a genetic disease, or a disease associated with a bacterial infection, or a predisposition to such a disease, and also is useful for determining identity, heredity or compatibility. Additional processes disclosed herein also are useful for such a diagnosis, for example, by determining the sequence of the target biological macromolecule obtained from the individual or by comparison of the target biological macromolecule with a corresponding known biological macromolecule.

The disclosed processes using IR-MALDI are suitable to analyzing more than one sample of biological macromolecule, particularly a large number of samples, for example, by depositing a plurality of compositions, each containing one or more biological macromolecules, on a solid support such as a chip, in the form of an array, if desired. In addition, the disclosed processes are suitable for multiplex analysis of a plurality of biological macromolecules contained in a one or a few compositions containing a liquid matrix. Each biological macromolecule in a plurality can be differentially mass modified, for example, to facilitate multiplex analysis. Accordingly, the processes are readily adaptable to high throughput assay formats.

A biological macromolecule particularly suitable for analysis by a process of IR-MALDI can be a nucleic acid, a polypeptide, a carbohydrate, or a proteoglycan, or can be a macromolecular complex such as a protein-protein complex or a nucleoprotein complex. For analysis, a target biological macromolecule can be immobilized to a substrate, particularly a solid support, which can be, for example, a bead, a flat surface, a chip, a capillary, a pin, a comb, or a wafer, and can be any of various materials, including a metal, a ceramic, a plastic, a resin, a gel, and a membrane. For example, the solid support can be a silicon wafer or a stainless steel flat surface. Since the processes as disclosed herein are particularly useful for analyzing a large number of target biological macromolecules in high throughput assays, it can be particularly useful to immobilize a plurality of target biological macromolecules in

an array on a solid support. Immobilization can be through a reversible linkage such as a photocleavable bond or a thiol linkage or a hydrogen bond, and the linkage can be cleaved using, for example, a chemical process, an enzymatic process, or a physical process, including during the mass spectrometric analysis procedure.

Where a target biological macromolecule is a nucleic acid, for example, the target nucleic acid can be immobilized by hybridization (hydrogen bonding) between a complementary capture nucleic acid molecule, which is immobilized to the solid support, and a portion of the nucleic acid molecule containing the target nucleic acid. It should be recognized, however, that, for some processes disclosed herein, at least a portion of the sequence containing the target nucleic acid should be distinct from the hybridizing portion of the target nucleic acid when immobilization is through hybridization to a capture nucleic acid, for example, where a detector oligonucleotide is to be hybridized to a sequence of the target nucleic acid.

Where the target biological macromolecule is a polypeptide, it can be immobilized to a solid support by binding to a reagent, which is conjugated to the solid support and specifically interacts with at least a portion of the target polypeptide or with a tag attached to the target polypeptide. Such a reagent can be, for example, an antibody that binds an epitope of the target polypeptide, or can be, for example, nickel ion, which binds to a polyhistidine sequence tag contained in the target polypeptide. A tag peptide such as a polyhistidine tag can be incorporated conveniently into a target polypeptide that is produced, for example, by an *in vitro* transcription or translation method.

A biological macromolecule to be analyzed can be conditioned prior to IR-MALDI mass spectrometric analysis. Conditioning improves the ability to analyze a particular biological macromolecule by IR-MALDI mass spectrometry, for example, by improving the resolution of the mass spectrum. If desired, the biological macromolecule can be isolated prior to conditioning or prior to mass spectrometric analysis.

A target biological macromolecule can be conditioned, for example, by ion exchange, by contact with an alkylating agent or a trialkylsilyl chloride, or

by incorporating at least one mass modified subunit into the biological macromolecule. For example, where the biological macromolecule is a nucleic acid, the target nucleic acid can be conditioned by phosphodiester backbone modification such as by cation exchange; by incorporating at least one nucleotide such as an N7-deazapurine nucleotide, an N9-deazapurine nucleotide, or a 2'-fluoro-2'-deoxynucleotide, each of which can reduce sensitivity of a nucleic acid to depurination; by incorporation of at least one mass modified nucleotide; or by hybridization of a tag probe to a portion of a nucleic acid molecule containing the target nucleic acid (see U.S. Patent No. 5,547,835).

A process for determining the identity of each target biological macromolecule in a plurality of target biological macromolecules can be performed, for example, by preparing a composition containing a plurality of differentially mass modified target biological macromolecules and a liquid matrix, which absorbs infrared radiation; determining the molecular mass of each differentially mass modified target biological macromolecule in the plurality by IR-MALDI mass spectrometry; and comparing the molecular mass of each differentially mass modified target biological macromolecule in the plurality with the molecular mass of a corresponding known biological macromolecule or fragment thereof. Where such a process is performed using a plurality of target biological macromolecules that are fragments of a biological macromolecule, the fragments can be prepared by contacting the biological macromolecules with at least one fragmenting agent that cleaves a bond involved in the formation of the biological macromolecules, particularly a bond between monomeric subunits of the biological macromolecule, to produce the fragment target biological macromolecules.

A target nucleic acid to be analyzed by IR-MALDI mass spectrometry can be in a biological sample and, if desired, can be amplified prior to analysis, then analyzed directly by IR-MALDI mass spectrometry. Alternatively, the amplified nucleic acid molecules can be contacted with a detector oligonucleotide, which can hybridize to a target nucleic acid sequence present in an amplified nucleic acid; a composition for IR-MALDI can be prepared by mixing the product of the reaction with a liquid matrix, which absorbs infrared radiation; and IR-MALDI

mass spectrometry can be performed. Detection of duplex nucleic acid molecules, which form by hybridization of the detector oligonucleotide and an amplified target nucleic acid, identifies the presence of the target nucleic acid in the biological sample.

Amplification of nucleic acid molecules, including a target nucleic acid molecule, can be performed using well known methods and commercially available kits. Amplification can utilize a polymerase, which can be a thermostable polymerase, such as Taq DNA polymerase, AmpliTaq FS DNA polymerase, Deep Vent (exo-) DNA polymerase, Vent DNA polymerase, Vent (exo$^-$) DNA polymerase, Vent DNA polymerase, Vent (exo$^-$) DNA polymerase, Deep Vent DNA polymerase, Thermo Sequenase, exo(-) *Pseudococcus furiosus* (*Pfu*) DNA polymerase, AmpliTaq, Ultman, 9 degree Nm, Tth, Hot Tub, *Pyrococcus furiosus* (*Pfu*) or *Pyrococcus woesei* (*Pwo*) DNA polymerase. Amplification processes include the polymerase chain reaction (Newton and Graham, *PCR* (BIOS Publ. 1994)); nucleic acid sequence based amplification; transcription-based amplification system, self-sustained sequence replication; Q-beta replicase based amplification; ligation amplification reaction; ligase chain reaction (Wiedmann et al., PCR Meth. Appl. 3:57-64 (1994); Barany, Proc. Natl. Acad. Sci., *USA* 88, 189-93 (1991)); strand displacement amplification (Walker et al., Nucl. Acids Res. 22:2670-77 (1994)); and variations of these methods, including, for example, reverse transcription PCR (RT-PCR; Higuchi et al., Bio/Technology 11:1026-1030 (1993)), and allele-specific amplification.

Where a nucleotide sequence of the target nucleic acid is amplified by PCR, well known reaction conditions are used. The minimal components of an amplification reaction include a template DNA molecule; a forward primer and a reverse primer, each of which is capable of hybridizing to the template DNA molecule or a nucleotide sequence linked thereto; each of the four different nucleoside triphosphates or appropriate analogs thereof; an agent for polymerization such as DNA polymerase; and a buffer having the appropriate pH, ionic strength, cofactors, and the like. Generally, about 25 to 30 amplification cycles, each including a denaturation step, an annealing step and an extension step, are performed, but fewer cycles can be sufficient or more

cycles can be required depending, for example, on the amount of the template DNA molecules present in the reaction. Examples of PCR reaction conditions are described in U.S. Patent No. 5,604,099.

A nucleic acid sequence can be amplified using PCR as described in U.S. Patent No. 5,545,539, which provides an improvement of the basic procedure for amplifying a target nucleotide sequence by including an effective amount of a glycine-based osmolyte in the amplification reaction mixture. The use of a glycine-based osmolyte improves amplification of sequences rich in G and C residues and, therefore, can be useful, for example, to amplify trinucleotide repeat sequences such as those associated with Fragile X syndrome (CGG repeats) and myotonic dystrophy (CTG repeats).

The presence of a target nucleic acid sequence in a biological sample also can be identified by specifically digesting nucleic acid molecules, which can be amplified nucleic acid molecules, containing the target nucleic acid with at least one appropriate nuclease; hybridizing the digested nucleic acid fragments with complementary capture nucleic acid sequences, which are immobilized on a solid support and can hybridize to a digested fragment of a target nucleic acid; preparing a composition for IR-MALDI, containing the immobilized fragments and a liquid matrix, which absorbs infrared radiation; and identifying immobilized fragments by IR-MALDI mass spectrometry (see International Publs. WO 96/29431 and WO 98/20019). The detection of nucleic acid fragments that were immobilized by hybridization to the complementary capture nucleic acid sequences identifies the presence of the target nucleic acid sequence in the biological sample. Immobilization of the nucleic acid fragments can be reversed prior to performing IR-MALDI or as a consequence of IR-MALDI mass spectrometry, for example, due to cleavage of an IR cleavable linkage during IR-MALDI.

The presence of a target nucleic acid in a biological sample also can be identified by performing on nucleic acid molecules obtained from the biological sample, a first polymerase chain reaction using a first set of primers, which are capable of amplifying a portion of the nucleic acid containing the target nucleic acid; preparing a composition containing the first amplification product and a

liquid matrix, which absorbs infrared radiation; and detecting the first amplification product in the composition by IR-MALDI mass spectrometry, thereby detecting the presence of the target nucleic acid in the biological sample. Such a process can include, prior to performing IR-MALDI, a second polymerase chain reaction on the first amplification product using a second set of primers, which are capable of amplifying at least a portion of the first amplification product containing the target nucleic acid (International Publ. WO 98/20019).

Processes for determining the identity of a subunit in a biological macromolecule, for example, for detecting a mutation in a nucleotide sequence, also are provided. The identity of a target nucleotide can be determined by hybridizing a nucleic acid molecule containing the target nucleotide with a primer oligonucleotide that is complementary to the nucleic acid molecule at a site adjacent to the target nucleotide; contacting the hybridized nucleic acid molecule with a complete set of dideoxynucleosides or 3'-deoxynucleoside triphosphates and a DNA dependent DNA polymerase, so that only the dideoxynucleoside or 3'-deoxynucleoside triphosphate that is complementary to the target nucleotide is extended onto the primer; preparing a composition containing the extended primer and a liquid matrix, which absorbs infrared radiation; and detecting the extended primer in the composition by IR-MALDI mass spectrometry. The identity of the target nucleotide is determined based on the dideoxynucleoside or 3'-deoxynucleoside triphosphate present in the extended primer, as determined by IR-MALDI mass spectrometry.

The absence or presence of a mutation in a target nucleic acid sequence also can be determined by hybridizing a nucleic acid molecule containing the target nucleic acid sequence with at least one primer, which has 3' terminal base complementarity to the target nucleic acid sequence; contacting the hybridized nucleic acid with an appropriate polymerase enzyme and sequentially with one of the four nucleoside triphosphates; preparing a composition containing the reaction product and a liquid matrix, which absorbs infrared radiation; and detecting the product in the composition by IR-MALDI mass spectrometry. Based on the molecular weight of the product, the presence or

absence of a mutation next to the 3′ end of the primer in the target nucleic acid molecule can be determined (International PCT application No. WO 98/20019).

A mutation in a target nucleic acid molecule also can be detected by hybridizing the target nucleic acid molecule with an oligonucleotide probe, to produce a hybridized nucleic acid, wherein a mismatch is formed at the site of a mutation; contacting the hybridized nucleic acid with a single strand specific endonuclease; preparing a composition containing the reaction product and a liquid matrix, which absorbs infrared radiation; and analyzing the composition by IR-MALDI mass spectrometry. The oligonucleotide probe used in this process has the sequence expected in a normal (unmutated) nucleic acid sequence corresponding to the target nucleic acid. The detection by IR-MALDI mass spectrometry of more than one nucleic acid fragment in the composition indicates that a mismatch was present in the hybridization product formed between the target nucleic acid and the oligonucleotide probe and, therefore, that the target nucleic acid molecule contains a mutation (International Publ. WO 98/20019).

The absence or presence of a mutation in a target nucleic acid sequence also can be identified by performing at least one hybridization of a nucleic acid molecule containing the target nucleic acid sequence with a set of ligation educts and a DNA ligase; preparing a composition for IR-MALDI containing the reaction product and a liquid matrix, which absorbs infrared radiation; and analyzing the composition by IR-MALDI mass spectrometry. Using such a process, the detection of a ligation product in the composition identifies the absence of a mutation in the target nucleic acid sequence, whereas the detection only of the set of ligation educts in the composition identifies the presence of a mutation in the target nucleic sequence.

A process of detecting the presence of ligation product by IR-MALDI mass spectrometry, as disclosed above, also can detect the presence of a target nucleic acid by performing at least one hybridization on a nucleic acid molecule containing the target nucleic acid with a set of ligation educts and a thermostable DNA ligase; preparing a composition containing the reaction product and a liquid matrix, which absorbs infrared radiation; and identifying a

ligation product in the composition by IR-MALDI mass spectrometry. The formation of a ligation product indicates the presence of the target nucleic acid.

A process as disclosed herein also provides a means of using IR-MALDI mass spectrometry to determine the identity of a target polypeptide by comparing the masses of defined peptide fragments of the target polypeptide with the masses of corresponding peptide fragments of a corresponding known polypeptide. Such a process can be performed, for example, by obtaining the target polypeptide by *in vitro* translation, or by *in vitro* transcription followed by translation of a nucleic acid encoding the target polypeptide; contacting the translated polypeptide with at least one fragmenting agent that cleaves at least one peptide bond in the polypeptide; preparing a composition for IR-MALDI containing the peptide fragments and a liquid matrix, which absorbs IR radiation; determining the molecular mass of at least one of the peptide fragments by IR-MALDI mass spectrometry; and comparing the molecular mass of the peptide fragments with the molecular mass of peptide fragments of a corresponding known polypeptide. The masses of the peptide fragments of a corresponding known polypeptide either can be determined in a parallel reaction with the target polypeptide, wherein the corresponding known polypeptide also is contacted with the agent; can be compared with known masses for peptide fragments of a corresponding known polypeptide contacted with the particular cleaving agent; or can be obtained from a database of polypeptide sequence information using algorithms that determine the molecular mass of peptide fragment of a polypeptide. Such a process is particularly useful, for example, for identifying mutations and, therefore, for screening for certain genetic disorders, for example, a single base mutation that introduces a STOP codon into an open reading frame of a gene, since such a mutation results in premature protein truncation; or a change in the encoded amino acid in an allelic variant of a polymorphic gene, for example, a single base change that results in an amino acid change of alanine to glycine, since polypeptides containing the different amino acids can be distinguished based on their masses.

A process of using IR-MALDI to analyze a target polypeptide to obtain information regarding the encoding nucleic acid can be used for identifying the presence of nucleotide repeats, particularly an abnormal number of nucleotide repeats, by determining the identity of a target polypeptide encoded by such repeats. An abnormal number of nucleotide repeats can be identified by using IR-MALDI mass spectrometry to compare the mass of a target polypeptide with that of a corresponding known polypeptide.

A target polypeptide can be obtained by translating an RNA molecule encoding the target polypeptide *in vitro*. If desired, the RNA molecule can be obtained by *in vitro* transcription of a nucleic acid encoding the target polypeptide. Translation of a target polypeptide can be effected by directly introducing an RNA molecule encoding the polypeptide into an *in vitro* translation reaction or by introducing a DNA molecule encoding the polypeptide into an *in vitro* transcription/translation reaction or into an *in vitro* transcription reaction, then transferring the RNA to an *in vitro* translation reaction.

In vitro transcription and *in vitro* translation kits are well known in the art and commercially available. *In vitro* translation systems include eukaryotic cell lysates such as rabbit reticulocyte lysates, rabbit oocyte lysates, human cell lysates, insect cell lysates and wheat germ extracts. Such lysates and extracts are can be prepared or are commercially available (Promega Corp.; Stratagene, La Jolla CA; Amersham, Arlington Heights IL; and GIBCO/BRL, Grand Island NY). *In vitro* translation systems generally contain macromolecules such as enzymes; translation, initiation and elongation factors; chemical reagents; and ribosomes. Mixtures of purified translation factors, as well as combinations of lysates or lysates supplemented with purified translation factors such as initiation factor-1 (IF-1), IF-2, IF-3 (alpha or beta), elongation factor T (EF-Tu) or termination factors, also can be used for mRNA translation *in vitro*. If desired, incubation can be performed in a continuous manner, whereby reagents are flowed into the system and nascent polypeptides removed or left to accumulate, using a continuous flow system as described by Spirin et al. (Science 242:1162-64 (1988)). Such a process can be desirable for large scale production of nascent polypeptides.

An *in vitro* translation reaction using a reticulocyte lysate, for example, can be carried out by mixing ten μl of a reticulocyte lysate with spermidine, creatine phosphate, amino acids, HEPES buffer (pH 7.4), KCl, MgAc and the RNA to be translated, and incubated for an appropriate time, generally about one hour at 30°C. The optimum amount of MgAc for obtaining efficient translation varies from one reticulocyte lysate preparation to another and can be determined using a standard preparation of RNA and a concentration of MgAc up to about 1 mM. The optimal concentration of KCl also can vary depending on the specific reaction. For example, 70 mM KCl generally is optimal for translation of capped RNA, whereas 40 mM generally is optimal for translation of uncapped RNA.

A wheat germ extract can be prepared as described by Roberts and Paterson (Proc. Natl. Acad. Sci., *USA* 70:2330-2334 (1973)) and can be modified as described by Anderson (Meth. Enzymol. 101:635 (1983)), if desired. The protocol also can be modified according to manufacturing protocol L418 (Promega Corp.). Generally, wheat germ extract is prepared by grinding wheat germ in an extraction buffer, followed by centrifugation to remove cell debris. The supernatant is separated by chromatography from endogenous amino acids and from plant pigments that are inhibitory to translation. The extract also is treated with micrococcal nuclease to destroy endogenous mRNA, thereby reducing background translation to a minimum. The wheat germ extract contains the cellular components necessary for protein synthesis, including tRNA, rRNA and initiation, elongation and termination factors. The extract can be optimized further by the adding an energy generating system such as phosphocreatine kinase and phosphocreatine; MgAc is added at a level recommended for the translation of most mRNA species, generally about 6.0 to 7.5 mM magnesium (see, also, Erickson and Blobel Meth. Enzymol. 96:38 (1982)), and can be modified, for example, by adjusting the final ion concentrations to 2.6 mM magnesium and 140 mM potassium, and the composition to pH 7.5 (U.S. Patent No. 4,983,521). Translation in wheat germ extract also can be performed as described in U.S. Patent No. 5,492,817.

For determining the optimal *in vitro* translation conditions or the extent of the reaction, translation of mRNA in an *in vitro* system can be monitored, for example, by mass spectrometric analysis. Monitoring also can be performed, for example, by adding one or more radioactive amino acids such as ^{35}S-methionine and measuring incorporation of the radiolabel into the translation products by precipitating the proteins in the lysate such as with TCA and counting the amount of radioactivity present in the precipitate at various times during incubation. The translation products also can be analyzed by immunoprecipitation or by SDS-polyacrylamide gel electrophoresis (see, for example, Sambrook et al., *Molecular Cloning: A laboratory manual* (Cold Spring Harbor Laboratory Press 1989); Harlow and Lane, *Antibodies: A laboratory manual* (Cold Spring Harbor Laboratory Press 1988)). A labeled non-radioactive amino acid also can be incorporated into a nascent polypeptide. For example, the translation reaction can contain a mis-aminoacylated tRNA (U.S. Patent No. 5,643,722). A non-radioactive marker can be mis-aminoacylated to a tRNA molecule and the tRNA amino acid complex is added to the translation system. The system is incubated to incorporate the non-radioactive marker into the nascent polypeptide and polypeptides containing the marker can be detected using a detection method appropriate for the marker. Mis-aminoacylation of a tRNA molecule also can be used to add a marker to the polypeptide in order to facilitate isolation of the polypeptide. Such markers include, for example, biotin, streptavidin and derivatives thereof (U.S. Patent No. 5,643,722).

In vitro transcription and translation reactions also can be performed simultaneously using, for example, a commercially available system such as the Coupled Transcription/Translation System (Promega Corp, catalog # L4606, # 4610 or # 4950). Coupled transcription and translation systems using RNA polymerases and eukaryotic lysates are described in U.S. Patent No. 5,324,637. Coupled *in vitro* transcription and translation also can be carried out using a prokaryotic system such as a bacterial system, for example, *E. coli* S30 cell-free extracts (Zubay, Ann. Rev. Genet. 7:267 (1973)).

A target polypeptide also can be obtained from a host cell transformed with and expressing a nucleic acid encoding the target polypeptide. The nucleic

acid encoding the target polypeptide can be amplified, for example, by PCR, inserted into an expression vector, and the expression vector introduced into a host cell suitable for expressing the polypeptide encoded by the target nucleic acid. Host cells can be eukaryotic cells, particularly mammalian cells such as human cells, or prokaryotic cells, including, for example, *E. coli*. Eukaryotic and prokaryotic expression vectors are well known in the art and can be obtained from commercial sources. Following expression in the host cell, the target polypeptide can be isolated using methods as disclosed herein. For example, if the target polypeptide is fused to a polyhistidine tag peptide, the target polypeptide can be purified by affinity chromatography on a chelated nickel ion column.

A target polypeptide can be produced from an amplified nucleic acid encoding the target polypeptide. Where a target polypeptide is produced, for example, from an amplified nucleic acid, it can be useful to operably link one or more transcription or translation regulatory elements to the nucleic acid or encoded polypeptide. Thus, a forward or reverse PCR primer can contain, if desired, a nucleotide sequence of a promoter, for example, a bacteriophage promoter such as an SP6, T3 or T7 promoter. Amplification of a nucleic sequence using such a primer produces an amplified nucleic acid operably linked to the promoter, i.e., the promoter is situated in the amplified nucleic acid such that it performs the function of a promoter. Such a nucleic acid can be used in an *in vitro* transcription reaction to transcribe the amplified target nucleic acid sequence.

A primer, for example, the forward primer, also can contain regulatory sequence elements necessary for translation of an RNA in a prokaryotic or eukaryotic system. In particular, where it is desirable to perform a translation reaction in a prokaryotic translation system, a primer can contain an operably linked prokaryotic ribosome binding sequence (Shine-Dalgarno sequence), which is located downstream of a promoter sequence and about 5 to 10 nucleotides upstream of the initiation codon.

A primer also can contain an initiation (ATG) codon, or complement thereof, as appropriate, located downstream of a promoter, if present, such that

amplification of the target nucleic acid results in an amplified target sequence containing an operably linked ATG codon, which is in frame with the desired reading frame. The reading frame can be the natural reading frame or can be any other reading frame. Where the target polypeptide is not a naturally occurring polypeptide, operably linking an initiation codon to the nucleic acid encoding the target polypeptide allows translation of the target polypeptide in the desired reading frame.

A primer, generally the reverse primer, also can contain a sequence encoding a STOP codon in one or more of the reading frames, to assure proper termination of the target polypeptide. Further, by incorporating into the reverse primer sequences encoding three STOP codons, one into each of the three possible reading frames, optionally separated by several residues, additional mutations that occur downstream (3') of a mutation that otherwise results in premature termination of a polypeptide can be detected.

A forward or reverse primer also can contain a nucleotide sequence, or the complement of a nucleotide sequence (if present in the reverse primer), encoding a second polypeptide. The second polypeptide can be a tag peptide, which interacts specifically with a particular reagent, for example, an antibody. A second polypeptide also can have an unblocked and reactive amino terminus or carboxyl terminus.

The fusion of a tag peptide to a target polypeptide or other polypeptide of interest allows the detection and isolation of the polypeptide. A target polypeptide encoded by a nucleic acid linked in frame to a sequence encoding a tag peptide can be isolated from an *in vitro* translation reaction mixture using a reagent that interacts specifically with the tag peptide, then the isolated target polypeptide can be subjected to IR-MALDI mass spectrometry, as disclosed herein. It should be recognized that an isolated target polypeptide fused to a tag peptide or other second polypeptide is in a sufficiently purified form to allow IR-MALDI mass spectrometric analysis, since the mass of the tag peptide will be known and can be considered in the determination.

Numerous tag peptides and the nucleic acid sequences encoding such tag peptides, which aids in isolationg of anything linked thereto, generally

contained in a plasmid, are known and are commercially available (NOVAGEN). Any peptide can be used as a tag, provided a reagent such as an antibody that interacts specifically with the tag peptide is available or can be prepared and identified. Frequently used tag peptides include a myc epitope, which includes a 10 amino acid sequence from c-myc (see Ellison et al., J. Biol. Chem. 266:21150-21157 (1991)); the pFLAG system (International Biotechnologies, Inc.); the pEZZ-protein A system (Pharmacia); a 16 amino acid peptide portion of the *Haemophilus influenza* hemagglutinin protein; a GST polypeptide; and a polyhistidine peptide, which generally contains about four to twelve or more contiguous His residues, for example, His-6, which contains six His residues. Reagents that interact specifically with a tag peptide also are known in the art and are commercially available and include antibodies and various other molecules, depending on the tag, for example, metal ions such as nickel or cobalt ions, which interact specifically with a His-6 peptide; or glutathione, which can be conjugated to a solid support such as agarose and interacts specifically with GST.

A second polypeptide also can be designed to serve as a mass modifier of the target polypeptide encoded by the target nucleic acid. Accordingly, a target polypeptide can be mass modified by translating an RNA encoding the target polypeptide operably linked to a mass modifying amino acid sequence, where the mass modifying sequence can be at the amino terminus or the carboxyl terminus of the fusion polypeptide. Modification of the mass of the polypeptide derived from such a recombinant nucleic acid is useful, for example, when several polypeptides are analyzed in a single IR-MALDI mass spectrometric analysis, since mass modification can increase resolution of a mass spectrum and allow for analysis of two or more different target polypeptides by multiplexing.

Tagged peptides

Polypeptides can be modified by addition of a peptide or polypeptide fragment to the target polypeptide. For example, a target polypeptide can be modified by translating the target polypeptide to include additional amino acids, such as polyhistidine, polylysine or polyarginine. These modifications serve aid

in purification, identification, and immobilization (and also in IR mass spectrometry). Modifications can be added post-translationally or can be encoded by a recombinant nucleic acid containing a sequence of nucleics that encode the target polypeptide.

Where a plurality of target polypeptides is to be differentially mass modified, each target polypeptide in the plurality can be mass modified, for example, using a different polyhistidine sequence, for example, His-4, His-5, His-6, and so on. The use of such a mass modifying moiety provides the further advantage that the moiety acts as a tag peptide, which can be useful, for example, for isolating the target polypeptide attached thereto. Accordingly, the disclosed processes permit multiplexing to be performed on a plurality of polypeptides, and, therefore, are useful for determining the amino acid sequences of each of a plurality of polypeptides, particularly a plurality of target polypeptides.

Primers for amplification can be selected such that the amplification reaction produces a nucleic acid that, upon transcription and translation, results in a non-naturally occurring polypeptide, for example, a polypeptide encoded by an open reading frame that is not a reading frame encoding a naturally occurring polypeptide. Accordingly, by appropriate primer design, in particular, by including an initiation codon in the desired reading frame and, if present, downstream of a promoter in the primer, a polypeptide produced from a target nucleic acid can be encoded by one of the two non-coding frames of the nucleic acid. Such a method can be used to shift out of frame STOP codons, which prematurely truncate a protein and exclude relevant amino acids, or to make a polypeptide containing an amino acid repeat more soluble. Primers useful for effecting the modifications disclosed herein can be obtained from commercial sources or can be synthesized using, for example, the phosphotriester method (see Narang et al., Meth. Enzymol. 68:90 (1979); U.S. Patent No. 4,356,270; see, also U.S. Patent Nos. 5,547,835; 5,605,798; and 5,622,824).

A non-naturally occurring target polypeptide also can be encoded by a 5′ or 3′ non-coding region of an exonic region of a nucleic acid; by an intron; or by a regulatory element such as a promoter sequence that contains, in one of the